



Review Article

The Impact of AI-Powered Software on Second Language (L2) Writing: A Systematic Literature Review

Angela Andrea Pérez ROA¹ 
Shanty HALIM² 

Corresponding author: angela.perez@unad.edu.co

DOI: <https://doi.org/10.31963/rial.v2i2.4801>

Received: 18/06/2024

Revised: 28/07/2024

Accepted: 29/07/2024

ABSTRACT

The utilization of artificial intelligence (AI)-powered tools in second language (L2) writing has evolved over the last decade. This attracted second-language writers to evaluate and improve their writing. This study aims to contribute to the understanding of the current state of AI-powered software in L2 writing, identify gaps in the literature, and investigate areas for future research. In this systematic literature review (SLR), we categorize the typology of AI-powered tools and their impact on L2 writing performance, discuss L2 writers' perceptions, and provide an overview of how they mitigate challenges and limitations in utilizing writing-assisted tools. The results of this SRL will bring implications for writing teachers, L2 researchers, and developers of AI-powered writing tools in the field of second language writing. This study also suggests that L2 writers should be aware of ethical aspects and academic integrity in utilizing and integrating AI in L2 writing.

Keywords: *Artificial intelligence, L2 writing, typology, ethics, and AWE*

To cite this article: Roa, A. A. P., Halim, S. (2024). *The Impact of AI-Powered Software on Second Language (L2) Writing: A Systematic Literature Review*. Research and Innovation in Applied Linguistics [RIAL], Vol 2 (2), 138-158. <https://doi.org/10.31963/rial.v2i2.4801>

¹ Universidad Nacional Abierta y a Distancia, Bogota, COLOMBIA. Email: angela.perez@unad.edu.co

ORCID: <https://orcid.org/0000-0002-2910-2399>

² Politeknik Negeri Ujung Pandang, Makassar, INDONESIA. Email: shantynurul@poliupg.ac.id

ORCID: <https://orcid.org/0009-0003-6160-9661>



This is an open access article under the [Creative Commons Attribution-ShareAlike 4.0 International License \(CC-BY-SA\)](https://creativecommons.org/licenses/by-sa/4.0/).

INTRODUCTION

Artificial Intelligence (AI) has had a significant impact on second language (L2) writing practices, such as grammar checkers, automated writing evaluation (AWE) tools, web-based paraphrasing tools, and plagiarism checkers (Almusharraf & Alotaibi, 2022; Barrot, 2021; Dizon & Gayed, 2021; Fu et al., 2022; Koltovskaia, 2022; Yan, 2023). They have become essential tools for L2 writers to improve their writing performance and productivity, thus assisting them in identifying and correcting errors, improving coherence and cohesion, enhancing vocabulary, and providing instant feedback on writing performance (Li et al., 2017; Nazari et al., 2021). However, little research has been addressed to investigate how AI-powered software The objective of this SRL was to conduct a systematic review of the impact of AI-powered software on L2 writing. It seeks to investigate and synthesize the existing literature on this topic to gain insights into the effects, perceptions, challenges, and limitations of AI-powered software in L2 writing practices.

Essentially, the advancement of knowledge is built on the basis of existing research. To expand frontiers of knowledge, it is essential to understand the current state of the field. This can be achieved through a comprehensive review of the relevant literature, which allows us to grasp the breadth and depth of existing research and identify areas that require further investigation (Xiao & Watson, 2019). With this in mind, this study contributes to the understanding of how AI-powered software shapes praxis and challenges in L2 writing. The results of this review will have implications for writing instructors, L2 researchers, and practitioners in the area of second language writing as well as for developers and designers of AI-powered software for language learning and writing support.

The ethics of using AI tool should become a serious attention to prevent learning lost and over-reliance on AI

METHOD

Research questions and validation

The research methodology for this study involved three stages: 1) preparing the review, 2) carrying out the review, and presenting the review (Xiao & Watson, 2019). The first process of this SLR was to formulate and pre-review the research questions to focus on the research objectives. Research inquiries were formed, constructed, and validated to ensure the reliability and validity of the review outcomes (Shaffril et al., 2021; Wee & Banister, 2015). In this study, the pre-review process of the research questions involved a panel discussion with the research team members. This led to the formulation of three research questions that lead the study:

- 1) What are the typologies of AI-powered writing tools, and how do they affect L2 writing?
- 2) How do L2 writers perceive AI-powered writing tools in their writing practice?
- 3) What are the challenges and limitations of the AI-powered writing software in supporting L2 writing?

These research questions were synthesized and merged from six initial questions to ensure that they collectively addressed the key aspects of the study objective and provided a comprehensive framework for the SLR process. This process involved a systematic and rigorous review, discussion, and refinement by the research team, ensuring that the final research questions were well aligned with the research objectives and scope of the review. Understanding the various types of these tools and their specific effects is crucial for identifying the most effective ones and how they enhance or hinder the writing process. Additionally, capturing L2 writers' perceptions helps determine the acceptance, usability, and practical implications of these tools, providing valuable insights into user preferences and challenges. Lastly, identifying the challenges and limitations of AI-powered writing tools is essential for guiding future improvements and ensuring they effectively support L2 writers.

Table 1.

SLR process (adapted from Xiao & Watson, (2019))

SLR Stages	Action Plan	Description
Preparing the review	Problem formulation	- Constructing research inquiries - Initial mapping
	Establishing the review protocol	- Creating the review procedures - Evaluating the SLR procedures
Carrying out the review	Conducting the literature search	- Select the scientific databases for the literature search (e.g., Google scholar, Crossref, and Scopus) - Start with a small set of focused keywords, synonyms, and related terms - Use Publish or Perish (PoP)
	Screen the articles	- Inclusion criteria - Screening procedure
	Asses the quality	- Criteria for quality assessment - Quality assessment procedure
	Extract the data	- Coding the papers (multi raters or coders) - Review the entire paper
	Analyze and synthesize the data	- Organize the coded papers - Synthesize the data and visualize them through tables, charts, or figures.
	Reporting the review	Report findings

Literature search

The researchers used “Publish or Perish” application to search the literature, including Scopus, Google Scholar, and Crossref databases. The search began with a small set of focused keywords, synonyms, and related terms that are directly related to the research questions, such as “Artificial Intelligence, or AI-writing software, or second language writing, or L2 writing tools, or automatic writing evaluation (AWE), or writing feedback”, and ChatGPT in L2 writing. Table 2 below shows the literature search results, including total search records and potentially relevant papers from the three largest scientific databases.

Table 2.
Literature search results

Search keywords	Total records			Potentially relevant papers		
	Scopus	Google Scholar	Crossref	Scopus	Google Scholar	Crossref
Artificial Intelligence, or AI-writing software	40*	503*	1000*	13	29	85
Second language writing, or L2 writing tools	149*	997*	1000*	33	22	13
Automatic writing evaluation (AWE), writing feedback	9*	989*	1000*	5	173	73
ChatGPT in L2 writing	28*	92*	1000*	15	27	14
	226	2.581	4.000	66	251	185

(*source: Publish or Perish)

Article screening

The screening method is contingent on the inclusion and exclusion criteria to prevent the acquisition of irrelevant articles and confirm that only high-quality, peer-reviewed research is considered for SRL. This strict approach allowed the review results to remain accurate and reliable, ensuring that the selected articles were in line with the research goals and enhanced the overall strength of the study. The quality of the review results depends on the sources or databases from which the articles are generated-“garbage-in, garbage-out”(Kilkenny & Robinson, 2018; Xiao & Watson, 2019)

Table 3.
Criteria for article inclusion and exclusion

Review criteria	Criteria for inclusion	Criteria for exclusion
Publication year	- From 2010 to 2023	- Published prior to 2010
Language	- Written in English	- Other than English
Subject	- Relevant to AI integration in L2 writing - Relevant to the use of writing-assisted tool, writing robot, writing evaluation, and automatic writing feedback in L2 writing	- Irrelevant to AI integration in L2 writing - Irrelevant to L2 writing discipline
Article type	- Original article - Research-based articles - Peer-reviewed	- Review articles, books, book reviews, proceedings, online magazines, reports - Not peer-reviewed
Journals/Publishers	- Papers are published by top tier journals, publishers, professional organizations, associations, and higher education press.	- Bogus journals/publishers - Listed in Beal’s list and predatoryjournal.com - Fake journal metric

Table 3 displays the criteria for including and excluding the sources used in the SLR. The inclusion criteria were articles that were directly related to artificial intelligence (AI) in the context of L2 writing, as well as those that were relevant to writing-assisted tools, writing robots, writing evaluation, and automatic writing feedback in L2 writing. Moreover, only

original research-based articles that underwent peer review and were published between 2010 and 2023 were considered. Additionally, articles written in English and published in reputable journals, publishers, professional organizations, associations, and the higher education press were included. On the other hand, the exclusion criteria included review articles, book reviews, proceedings, online magazines, and reports, as well as articles that were not peer-reviewed, published prior to 2010, written in languages other than English, published in bogus journals/publishers, listed in Beal's list or predatoryjournal.com, or with fake journal metrics.

Furthermore, the screening process yielded 6.807 articles (including duplicates) that met the inclusion criteria, while 12 non-English articles were excluded. Further refinement was achieved by applying the screening procedures, resulting in 502 relevant articles to AI in L2 writing, including their relevance to writing-assisted tools, writing robots, writing evaluation, automatic writing feedback in L2 writing, ChatGPT in L2 writing. Subsequently, the article type criterion narrowed down the selection to 371 articles. At this stage, the articles selection becomes more stringent, hundreds of articles were removed, including irrelevance to AI and L2 writing (n=94), book/book review (n=23), proceedings (n=29), online magazine (n=7), unreliable journal/publishers (n=27), not research-based/ review articles (n=24), and duplicates (n=33). Finally, the number of studies that merit for quality assessment were 134 articles (papers from iteration process were not included yet). These articles were then reviewed to ensure their qualities to merit for further analysis.

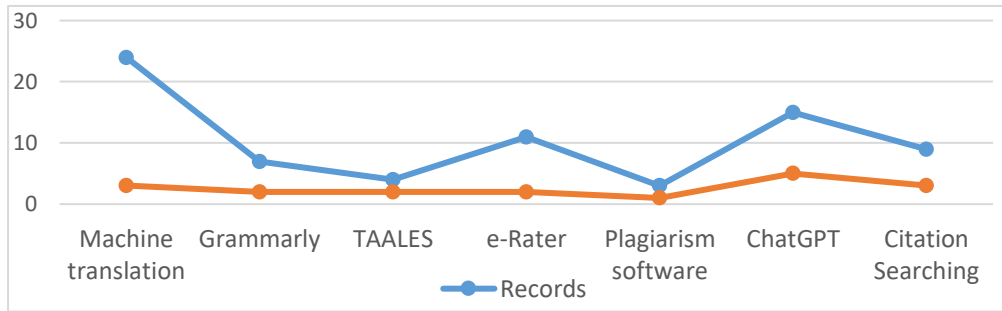
Quality assessment

One-hundred and thirty-four articles yielded from the screening process that merit for further quality assessment were retrieved and distributed to all the reviewers. Four reviewers, two researchers and two independent reviewers, were involved in the quality assessment process. They were assigned to qualitatively assess and categorize the papers into three levels: low, moderate, and high level quality (Petticrew & Roberts, 2006). After four weeks parallel independent quality assessment, the process was ended with a panel discussion involving all the reviewers to resolve any disagreements. As a result, the reviewers agreed to include 114 articles and exclude 20 articles due to the content quality, research method, context, and their implications to L2 writing practices.

Iterations

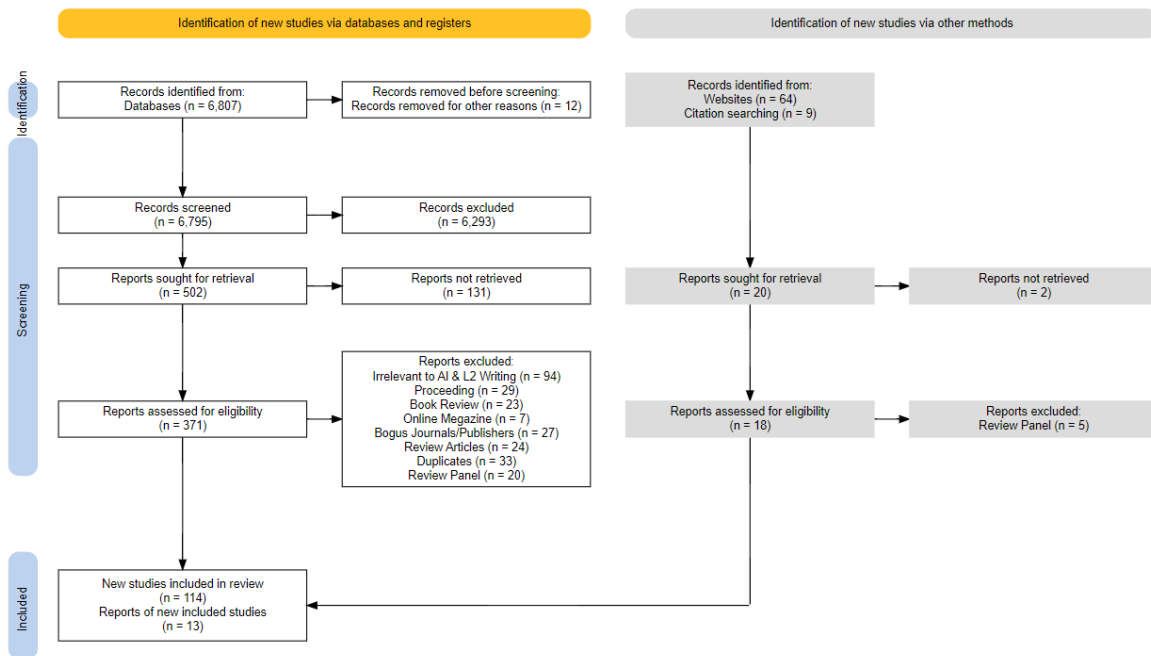
We also refined the keywords iteratively during the literature search to capture the most relevant literature. Iterations are essential to reduce workload in the SLR process (Lavallée et al., 2014; Xiao & Watson, 2019). We then continued with forward and backward search by doing a “citation chasing” and keywords refinement to track the citation trail that have influenced the current body of literature. The iteration process was also stringent in that the literature search

Figure 1.
Iterations



The figure shows the results of an iterative keyword search process, where different keywords were used in multiple iterations to search for relevant records. It includes five keywords: "Machine translation", "Grammarly", "TAALES", "e-Rater", "plagiarism software", and "ChatGPT". In the initial search, these keywords resulted in a total of 73 records. However, after a selection process, only 18 records were included for further analysis and review. The number of records selected for inclusion varied for each keyword, with "Machine translation" (n=3), "Grammarly" (n=2), "TAALES" (n=2), "e-Rater" (n=2), "Plagiarism software" (n=1), "ChatGPT" (n=5), and reference search (n=3). Finally, the review panel reviewed these selected articles and excluded five more articles due to the lack of research method rigor, so the number of remaining articles is thirteen.

Figure 2.
Screening for inclusion results (generated from PRISMA, Haddaway et al., 2022)



Extracting, analyzing and synthesizing the data

At this stage, the process involved several steps of analysis, including data extraction, analysis, synthesis, interpretation, reporting (Petticrew & Roberts, 2006). Data extraction involved extracting relevant information from the selected articles. As we worked as a team, we coded three papers together before splitting the tasks for individual review. It aims to ensure everyone involved in the process has the same procedure for coding the articles (Galvan, 2017). This process took four weeks to read, code, analyze, and synthesize the entire articles before conducting a panel discussion to resolve any disagreements and misunderstanding of the review outcomes. Finally, we present the results in the following sections.

FINDINGS

Typology of AI-powered software in L2 writing

The typology of AI-powered software for L2 (second language) writing were categorized based on their functionalities and purposes. The types of AI-powered software for L2 writing include machine translation tools, grammar and spell-checking tools, writing evaluation tools, writing assistance tools, and plagiarism detection tools. These tools utilize artificial intelligence algorithms and technologies to automate and assist various aspects of L2 writing, such as generating translations, providing grammar and spell-check suggestions, evaluating writing quality, offering writing suggestions, and detecting potential plagiarism.

Figure 3.

Typology of AI-enabled tools in L2 writing

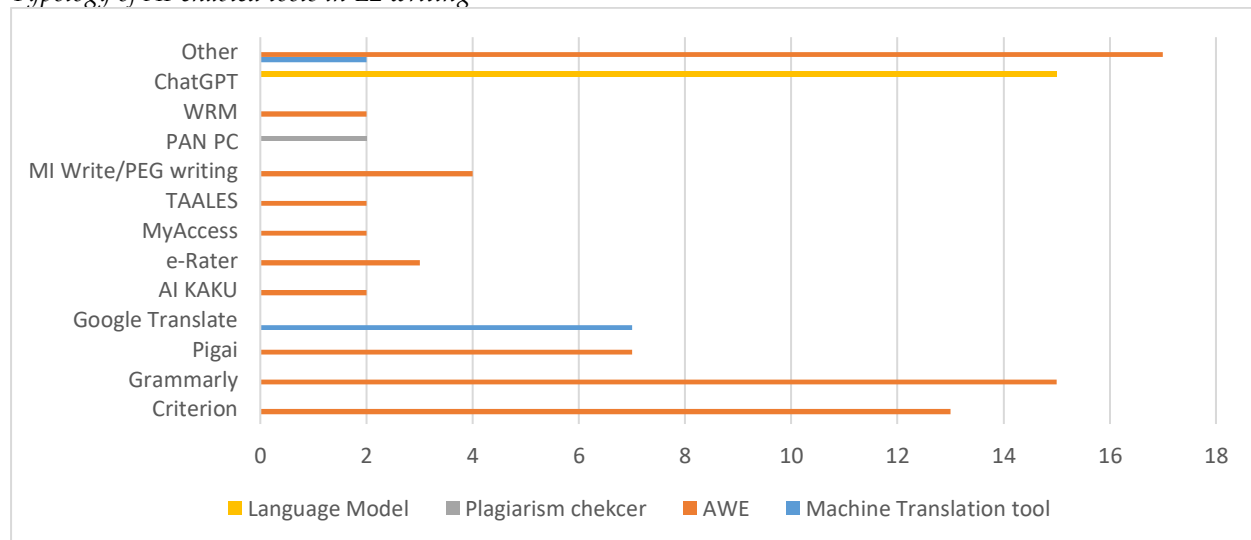


Figure 3 shows the number of studies conducted for different categories of language and text analysis tools, specifically focusing on machine translation, AWE (Automatic Writing Evaluation), and plagiarism checking. Grammarly, Criterion, and Pigai are the top three AWE tools that were extensively researched in L2 writing context, highlighting their

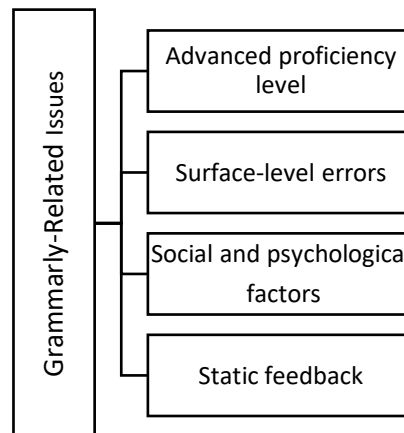
popularity and effectiveness in enhancing and supporting L2 writing. In the machine translation category, Google Translate stands out with the most favorable AI application used in L2 writing. On the other hand, AI KAKU, e-Rater, MyAccess, TAALES, MI Write/PEG writing, and WRM have fewer studies conducted, indicating a lack of academic research specifically targeting these AWE tools. More interestingly, ChatGPT has gained more attention in L2 writing since its launch in 2022. Its users are now growing very rapidly legitimating its ability to generate instant access to a vast repository of information.

Grammarly

The investigation reveals several insights regarding the effectiveness of Grammarly as an AI-powered assisted writing tool in L2 writing. First, it showed positive effects on non-native postgraduate students' cognitive, non-cognitive, and emotional domains of learning in L2 writing. It contributes to the overall improvement in these areas, suggesting its potential as a valuable tool in supporting L2 learners' writing skills development. However, it is important to note that Grammarly's benefits may be more pronounced for students with advanced language proficiency than for those with low proficiency. This indicates a potential limitation and highlights the need for targeted support for students with lower language skills.

Figure 4.

Grammarly-related issues in L2 writing



The findings also indicate that Grammarly has a significant impact on improving students' writing performance. The tool proved particularly effective in reducing errors in vocabulary usage, language use, and writing mechanics. This suggests that this can be a valuable resource for L2 writers to improve their overall accuracy and fluency. However, despite its benefits, students in these studies expressed a preference for peer feedback over Grammarly. They considered their peers as their true audience and placed a higher value on their feedback than on Grammarly's feedback. This highlights the importance of social and psychological factors in students' feedback engagement, and suggests that AWE tools such as Grammarly should be designed to complement and enhance rather than replace the role of peer feedback in writing instruction. Additionally, the findings also indicate that this

AWE tool is particularly effective in addressing local surface-level errors, such as articles, prepositions, and verb-noun agreements. This suggests that its strength lies in its ability to provide real-time corrective feedback for specific grammatical aspects. Moreover, using Grammarly results in fewer grammatical errors and increased lexical variation, which can be especially beneficial for novice L2 writers who struggle with effective language use.

Finally, the findings highlighted the importance of building trust with users and promoting a learning-oriented approach to feedback engagement in the design and use of AWE tools such as Grammarly in L2 classroom settings. Ensuring accuracy and clarity in error-flagging, providing specific feedback, and considering the social and psychological aspects of feedback reception are key factors to consider in the design and implementation of AWE tools to maximize their benefits. Although the findings provide valuable insights into the benefits and effectiveness of Grammarly, there are some gaps that could be addressed in future research. These include further investigation into the specific cognitive, non-cognitive, and emotional impacts of Grammarly on L2 learners' writing abilities, exploring the limitations or challenges faced by students with low language proficiency when using Grammarly, and understanding the long-term effects of using AWE tools such as Grammarly on students' writing skills beyond immediate performance improvements. Additionally, examining instructors' perceptions and experiences regarding the integration of AWE tools in L2 classrooms would provide insights into pedagogical practices and strategies for effective implementation.

Criterion

Based on the findings from the studies conducted on Criterion, several interesting issues emerge, which should be of concern to researchers: 1) AWE as complement to teachers' feedback, 2) discrepancies between teacher and Criterion feedback, 3) effectiveness of Criterion feedback, 4) individual differences and contextual factors in using Criterion, 5) teacher agency and cognition, scaffolding and instructional support, and 6) multiple attempts in using Criterion. Researchers should delve deeper into understanding the dynamics between teacher feedback and AWE, exploring how teachers can leverage AWE to enhance their feedback practices and support students' writing development effectively. The studies also emphasize the importance of considering individual differences, such as language proficiency levels and attitudes, when implementing Criterion. Therefore, researchers should delve deeper into understanding how these factors influence students' experiences with AWE systems and the outcomes of using Criterion. Contextual factors, such as the writing tasks, instructional practices, and learner characteristics, should also be examined to determine how they interact with the use of Criterion.

In terms of using Criterion as writing-assisted tool, L2 novice writers will need potent scaffolding strategies to help them use the tool effectively and improve their metalinguistic ability. Metalinguistic competence is central to understanding the AWE feedback to mitigate

the linguistics problems during the writing process. Therefore, teachers must be able to expose lower-performing students to multiple attempts using the tool, thus improving their skill development, error identification, self-regulated learning, self-efficacy, confidence, exposure to different writing tasks, and familiarization with the AWE system. Most of the studies also emphasize that teachers' feedback provides meaningful experience over the use of AWE tools, including Criterion. Beginning writers should be mentored in using AWE so they can interpret the feedback from Criterion meaningfully and navigate their writing skills using the tool effectively.

Pigai

Pigai, another popular AWE tool in China, is an online scoring and feedback system that is primarily used for English writing assessment. It is designed to evaluate students' essays based on predefined criteria and provide feedback on grammar, vocabulary, coherence, and other aspects of writing. It aims to improve students' English writing level by providing instant feedback and reducing the workload of teachers. In other words, it focuses on assessing and improving students' writing skills in the English language. Despite the valuable insights provided by the studies, there are some gaps that need further exploration. Firstly, there is a need for more diverse and large-scale studies to generalize the findings across different contexts and populations. Secondly, more research is required to understand the factors that shape students' attitudes and perceptions towards AWE feedback, as well as their appropriation of such feedback. Additionally, while some studies highlight the benefits of automated feedback tool, further investigation is needed to understand the optimal use and effectiveness of these different feedback types. Lastly, there is a need for more comprehensive research on the role of engagement as a mediating variable in the use of feedback and its impact on students' writing development.

The findings also suggest that learners' intention to use AWE feedback is influenced by various factors, including perceived usefulness, attitude towards use, computer self-efficacy, and perceived ease of use. Pigai systems have the potential to improve students' writing skills, reduce the workload of teachers, and enhance students' initiative and writing level. However, the effectiveness of AWE feedback depends on the engagement of students and the specific feedback features provided. While teacher feedback tends to address more error categories and provide corrections, AWE feedback highlights errors without specific corrections. The findings also emphasize the importance of engagement and individual differences in the use of feedback and its impact on writing development.

The study provides valuable insights into the factors that influence students' acceptance of AWE feedback, offering guidance for developers and practitioners to create more effective systems. Key determinants such as perceived usefulness, attitude towards use, computer self-efficacy, and perceived ease of use significantly shape students' intention to use the tool. However, the study's generalizability is limited to Chinese college students

using the Pigai AWE tool. Challenges include interpreting submission frequency as a measure of engagement and comparing the effectiveness of teacher feedback and AWE feedback due to variations in assignments and students' English proficiency levels. Students show a preference for Pigai as automated corrective feedback (ACF) in improving sentence writing and vocabulary, but there is less focus on higher-level revisions.

Further research is needed to explore the broader impact of AWE on writing revision, address specific challenges in College English writing, and expand understanding of AWE adoption in different L2 writing contexts. The study contributes invaluable insights into the factors that shape students' acceptance of AWE feedback, while simultaneously acknowledging the contextual limitations of the research. Further investigation is warranted to explore the impact of AWE features on broader aspects of writing revision, to devise effective pedagogical strategies that address the specific challenges encountered in College English writing, and to expand the understanding of AWE adoption across diverse L2 writing contexts, thereby yielding broader implications for educational practice.

Google Translate

The findings suggest that Google Translate (GT) has brought a positive impact on learners' vocabulary, grammar, and word-spelling development. They demonstrated improvements in these areas, including enhanced accuracy, conciseness, and coherence of their writings. The tool is also perceived helpful for beginners in improving communication by overcoming language barriers. For example, it assists students to improve fluency, cohesion, and the production of syntactically complex sentences. However, there was a tendency for learners to produce inaccurate and literal translations at times, indicating the need for careful evaluation of the tool's output. While the tool primarily focused on vocabulary, grammar, and spelling improvement, its benefits in promoting syntactic and lexical complexity were less evident. Overall, the tool played a valuable role in supporting language development, but further refinement and improvement are necessary to address its limitations and optimize its effectiveness. For example, L2 writers with low proficiency will need teachers' guidance to help them mitigate inaccuracies and errors during the writing process. In other words, GT helps low proficiency L2 writers to improve their grammar, vocabulary, and spelling mastery, but mostly unable to produce texts with lexico-grammatical complexities.

ChatGPT in L2 Writing

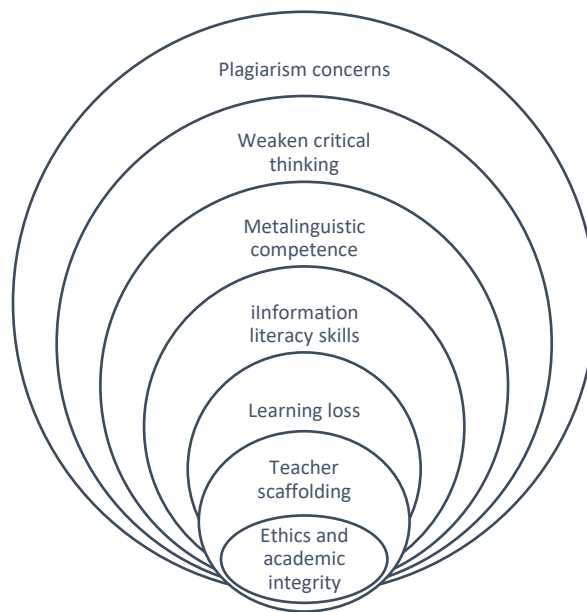
The review results highlight several interesting issues regarding the use of ChatGPT in L2 writing, offering benefits and opportunities for students to navigate their writing skills. However, the findings also showed its impact on writing ethics and academic integrity. In one study, researchers found that students with a high level of language learning motivation (LLM) exhibited greater satisfaction with their use of ChatGPT, particularly in terms of perceived helpfulness, appropriateness of style/tone, performance, overall satisfaction, and encouragement. These findings suggest that the ChatGPT can effectively enhance English

writing skills when used in conjunction with a platform such as RECIPE, which guides students to obtain satisfactory responses. However, no statistically significant difference was found in terms of trustworthiness and credibility, indicating the need for caution when using the ChatGPT as a source of information.

The articles also shed light on the benefits and limitations of ChatGPT in different contexts. While it can enhance language learning and improve efficiency in L2 writing, users must exercise caution to ensure the accuracy, credibility, and responsible usage of the tool. It is crucial that educators and researchers engage in discussions about the ethical implications of using AI technologies, such as ChatGPT, to ensure its responsible integration into various domains. However, cautious usage, critical evaluation, and ongoing improvements are essential to ensure the reliability, accuracy, and responsible integration of the ChatGPT in these contexts.

Figure 5.

ChatGPT-related issues in L2 writing



As synthesized from the selected articles, the utilization of ChatGPT in L2 writing is associated with several issues, including ethics and academic integrity, the need for teacher support and guidance, learning loss, the need for information literacy skills, lack of metalinguistic competence, weakened critical thinking, and plagiarism concerns. Almost all of the literature analyzed shows that there is concern that the misuse of ChatGPT has the potential to change students' writing behavior. Moreover, the ease of access offered by ChatGPT can trigger attitudes towards dishonesty, irresponsibility, and disrespect for the value of the original work. There is a serious concern about leveraging teachers' role in supporting and guiding students in navigating AI-powered tools to support their learning. In other words, there is a growing need to anticipate learning loss, weaken critical thinking,

and plagiarism. Thus, the utilization of the ChatGPT in L2 writing requires caution and calls for developing students' information literacy skills.

Effects on L2 writing accuracy

The effects of AI-powered software on the accuracy of L2 writing are multifaceted. On the one hand, these tools can potentially enhance the accuracy of L2 writing by providing real-time feedback, correcting errors, suggesting improvements, and assisting writers in generating accurate translations. They can also help learners improve their grammar, vocabulary, and writing skills through continuous feedback and guidance.

On the other hand, the accuracy of L2 writing facilitated by AI-powered software may also be influenced by certain factors. For instance, the accuracy of machine translation tools can depend on the complexity of the source text and the target language, and may not always produce completely accurate translations. Grammar and spell-checking tools may not catch all errors or may generate false positives. Writing evaluation tools may have limitations in accurately assessing the quality of writing beyond surface-level errors. Writing assistance tools mainly offer suggestions that are not always contextually appropriate or may not align with the writer's intentions. Plagiarism detection tools also generate false positives or false negatives, and do not always accurately identify instances of plagiarism. In other words, the effects of AI-powered software on the accuracy of L2 writing were influenced by the type of software used, its functionalities, and its limitations. It is important for L2 writers to be aware of the strengths and limitations of such tools and use them judiciously in conjunction with other language learning strategies to improve their overall writing accuracy.

L2 writers' perceptions of AI-powered software

The studies reveal that the use of AI-powered assist-writing tool (e.g., Grammarly) can have positive effects on various domains of learning in English academic writing for non-native English writers. However, it was noted that students with lower language proficiency could not benefit from Grammarly as much as those with advanced proficiency. The analysis of performance scores showed a significant improvement in writing quality after using it, particularly in terms of vocabulary usage, language use, and mechanics of writing. Additionally, students preferred peer feedback over Grammarly, valuing their peers' feedback as their true audience. Nevertheless, the study confirms the overall positive effects of AWE tools in enhancing writing skills and highlights the potential of real-time corrective feedback to support L2 writing, especially among novice writers.

The results also indicate that Grammarly was more effective in addressing local surface-level errors, such as articles, prepositions, and verb-noun agreement. When students wrote with the assistance of Grammarly, they demonstrated fewer grammatical errors and increased lexical variation. These findings emphasize the potential of predictive text and real-time feedback to support L2 writing, particularly for learners who may struggle with

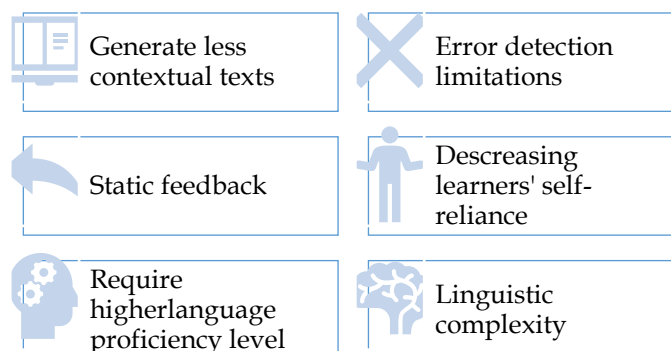
effective writing in their second language. The study further reveals that students successfully addressed a high percentage of Grammarly-flagged usages, highlighting the accuracy and effectiveness of error correction facilitated by the tool. These findings have implications for source-based academic writing instruction, the development of AWE tools, and the design of AWCF (Automated Writing Correction and Feedback) tools, emphasizing the importance of specific and accurate feedback, building trust with users, and promoting a learning-oriented approach to feedback engagement in L2 classroom settings.

Challenges and limitations of AI-powered software

The utilization of AI-powered software in L2 writing endeavors extends beyond the final written output, encompassing its auxiliary function as a supportive tool in the writing process. By assuming an assistive role, such software aids L2 writers in mitigating errors, thereby fostering the development of their metalinguistic awareness and lexico-grammatical competencies. Despite their potential benefits, they also face challenges and limitations, including students' language proficiency, linguistics complexity, feedback quality, error detection ability, static feedback, and the decrease of students' self-reliance.

Figure 5 presents several constraints associated with AI-powered writing software, which sparks researchers' inquisitiveness regarding the tool's inherent limitations and its inability to fulfill certain crucial functions. Firstly, novice writers with limited language proficiency encounter challenges in conducting linguistic analysis and comprehending feedback provided by AI-powered writing tools. The static nature of the tool's response restricts opportunities for users to confirm their understanding, resulting in uninteractive feedback. This kind of feedback lacks the inherent characteristics of human inquiry, which typically involves seeking explanations, support, justification, and reinforcement for acquired knowledge. In contrast, an AI machine operates as an artificial entity driven by an algorithmic system designed by its creators. Consequently, users of the tool still need to possess metalinguistic competence to effectively analyze the feedback generated by the AI system. Writers with low proficiency tend to accept the analysis results provided by the AI tool due to their limited linguistic skills in evaluating the feedback's accuracy and relevance.

Figure 5.
Challenges and limitations of AI-enabled writing tool



Language complexity is an inherent characteristic of language, encompassing its dynamic and intricate nature. AI-powered software may encounter challenges in comprehending the subtleties, nuances, idioms, and cultural references embedded within language, leading to potential inaccuracies in translations or suggestions. The intricate interplay of grammar, vocabulary, syntax, semantics, lexicon, idiomatic expressions, and cultural connotations poses difficulties for AI algorithms in capturing the complete meaning and essence of language. For instance, a phrase like "kick the bucket" to denote someone's demise may confound AI systems reliant on literal translation. Additionally, cultural references, humor, and contextual wordplay that hinge on shared knowledge may elude accurate interpretation by AI. The limitations of AI in capturing the multifaceted aspects of language highlight the challenge of reproducing the richness and complexity of human communication. Consequently, the tools may possibly generate less contextual texts and lead to misinterpretations.

On the other hand, grammar and spell-checking tools may not consistently detect all errors or could generate erroneous feedback, resulting in feedback or corrections that are not entirely accurate. These tools have limitations that can impact their effectiveness. One limitation is their potential to mistakenly flag correct language usage as incorrect, known as false positives, or to overlook actual errors, known as false negatives. These inaccuracies can lead to confusion and unnecessary revisions for the writer. Moreover, the use of AI-assisted tool may lead to a decrease in learners' self-reliance, critical thinking, and language skills development, as they may become overly dependent on the software for corrections and suggestions. They will have a minimum exposure to a process of meaning-making writing activities that shapes students' learning and skill development.

DISCUSSION

Writing is a complex activity and requires the writers' literacy, reading navigation, and critical thinking skills to construct a cohesive, concrete, concise, clear, and grammatically correct sentences (Deane et al., 2008). Among the many types of AI-powered tools in L2 writing (see Figure 3), ChatGPT has received a global attention from scholarly communities in just within a year since its launch. However, the growing popularity of ChatGPT also legitimizes the need for a critical examination of its limitations and ethical implications, particularly in academic and educational contexts (Hong, 2023).

The findings regarding Grammarly in L2 writing instruction reveal several insights. Firstly, Grammarly worked better in students with advanced language proficiency compared lower proficiency students. They require more personalized assistance to fully leverage the advantages of such AI tools. This finding highlights the importance of differentiated instruction to address the diverse needs of students (Bantis, 2008; Chapman & King, 2009). Therefore, the combination of the "process approach to writing" and "differentiated instruction approach" can enhance the development of L2 writing instruction (Ismail, 2019). Thus, using AI-powered tools and differentiated instruction altogether in L2

writing class can help students mitigate language barriers and improve their writing skills more effectively. Moreover, the absence of social and psychological aspects of the writing process with Grammarly confirms that AI cannot replace the role of teacher-student interaction (e.g., peer-feedback and teacher-feedback) but rather as a complement or tool in writing (Muna et al., 2023). AWE tools like Grammarly can be valuable aids, but they should be integrated into the writing process in a way that complements and enhances the role of peer feedback and peer interaction, ultimately contributing to more effective and comprehensive writing education.

Similar to Grammarly, Criterion should also be used as complementary resources to assist L2 writing teachers in providing feedback to students, confirming the importance of teachers' agency and cognition in AI-assisted writing tool (Li, 2021). In the case of L2 novice writers, scaffolding strategies are essential to help them utilize AWE tools effectively and enhance their metalinguistic competence (Liang, 2007). This metalinguistic competence is central to understanding and addressing linguistic issues during the writing process (Gutierrez, 2008). Furthermore, the studies highlight the importance of teachers mentoring beginning writers in using AWE tools. This mentorship is essential for helping students interpret AWE feedback meaningfully, navigate the AWE system, and develop their writing skills effectively (Barrot, 2021).

Another AI-powered assisted writing assessment tool is Pigai which is widely used in China. As commonly found in other AI writing robots, it mainly points out errors without providing detailed corrections, while teachers' feedback typically covers a wider range of error categories and offers specific corrections. As Zhang (2022) proposed a refinement for Pigai, writing instructors and AWE developers should collaborate to upgrade the systems. Although AI-powered AWEs (e.g., Grammarly, Criterion, and Pigai) are designed for self-regulated writing tools, students with low metalinguistic competence will find them complicated without teachers' scaffolding and mentoring.

Although Grammarly, Criterion, and Pigai are widely used as AWE in L2 context, ChatGPT penetrates with a more advanced language model that can generate texts and information in just a few seconds. However, L2 writers (mostly novice) might not be able to use the tool effectively due to several issues (see Figure 4). The writing students must possess a thorough understanding of the ethical aspect and academic integrity, demonstrating respect for the principles of originality in their work. Apart from receiving the benefits, on the other hand, they are also worried about academic honesty and fairness in writing (Yan, 2023). With this in mind, plagiarism in L2 writing context with ChatGPT can also be a serious concern (Jarrah et al., 2023). Therefore, there is a need for institutional policies, procedures, and workshop regarding the use of ChatGPT in educational context (Cotton et al., 2023). Moreover, providing a workshop on improving the writing students' information literacy skills and linguistics competence can help them develop a more holistic and effective

approach to L2 writing. With this in mind, the writing teachers will be able to anticipate the learning loss with ChatGPT, thus promoting the students critical thinking and scaffolding while working with the tool. In other words, L2 writing with ChatGPT for students with limited metalinguistic competence requires coaching and mentoring strategies to guide them in effectively using the tool and understanding the feedback it provides, ultimately enhancing their language proficiency and writing abilities.

Regarding the L2 writing accuracy, most AWE tools are perceived effective in providing surface-level feedback, especially for students with advanced language proficiency (Xu & Zhang, 2021). However, novice writers might not be able to understand the AWE feedback without detailed corrections and explanations of why such errors occurred. Therefore, they often prefer receiving peer feedback to using AWE (Ginting & Fithriani, 2022). Therefore, challenges in using AWE tools in L2 writing context, including language model ChatGPT, are multifaceted (see Figure 5). These challenges encompass issues related to proficiency levels, feedback comprehensibility, and the preference for peer interaction in the writing process, highlighting the need for a balanced and supportive approach to integrating AWE tools into L2 writing instruction.

CONCLUSION

The research findings in the realm of L2 writing with AI-powered software shed light on a multitude of critical aspects. These findings emphasize several constraints faced by the students when incorporating AI tools into the writing practice. Writing is a complex and a meaning-making process, thus calling for teachers' effective scaffolding strategies and active mentoring. These strategies are crucial to help learners, particularly novices, navigate AI-powered tools effectively and enhance their metalinguistic ability. The utilization of AI-powered tool in L2 writing highlights the need for comprehensive guidance and training, as well as teacher presence and involvement in the writing process. In this context, teachers must continue to adapt and evolve their instructional methods to leverage the advantages of AI while mitigating its ethical challenges. These findings not only deepen our understanding of the dynamic relationship between AI technology and L2 writing but also emphasize the importance of a holistic and ethical approach to writing instruction. As technology continues to advance, teachers must navigate the ever-evolving landscape of AI-powered tools, using these insights to create effective, ethical, and empowering learning experiences for their students.

ACKNOWLEDGMENT

We would like to express our sincere gratitude to the Language Center and the Center for Business English and Computer Assisted Language Learning (CBE-CALL) Research for initiating and mediating the international collaboration between the two scholars from Politeknik Negeri Ujung Pandang-Indonesia and Universidad Nacional Abierta y a Distancia-Colombia.



CONFLICT OF INTEREST

No conflict of interest reported by the authors


REFERENCES


- Almusharraf, N., & Alotaibi, H. (2022). An error-analysis study from an EFL writing context: Human and Automated and Automated Essay Scoring Approaches. *Technology, Knowledge and Learning*. <https://doi.org/10.1007/s10758-022-09592-z>
- Bantis, A. M. (2008). *Using task based writing instruction to provide differentiated instruction for English language learners* [University of Southern California]. <https://files.eric.ed.gov/fulltext/ED504698.pdf>
- Barrot, J. S. (2021). Using automated written corrective feedback in the writing classrooms: Effects on L2 writing accuracy. *Computer Assisted Language Learning*, 36(4), 584–607. <https://doi.org/10.1080/09588221.2021.1936071>
- Chapman, C., & King, R. (2009). *Differentiated instructional strategies for writing in the content areas* (2nd ed.). Corwin Press.
- Cotton, D. R. E., Cotton, P. A., & Shipway, J. R. (2023). Chatting and cheating: Ensuring academic integrity in the era of ChatGPT. *Innovations in Education and Teaching International, Advanced Publication*. <https://doi.org/10.1080/14703297.2023.2190148>
- Deane, P., Odendahl, N., Quinlan, T., Fowles, M., Welsh, C., & Bivens-Tatum, J. (2008). Cognitive Models of Writing: Writing Proficiency as a Complex Integrated Skill. In *ETS Research Report Series*. <https://doi.org/10.1002/j.2333-8504.2008.tb02141.x>
- Dizon, G., & Gayed, J. M. (2021). Examining the impact of Grammarly on the quality of mobile L2 writing. *The JALT CALL Journal*, 17(2), 74–92.
- Fu, Q.-K., Zou, D., Xie, H., & Cheng, G. (2022). A review of AWE feedback: types, learning outcomes, and implications. *Computer Assisted Language Learning, Advanced Online Publication*. <https://doi.org/https://doi.org/10.1080/09588221.2022.2033787>
- Galvan, J. L. (2017). *Writing Literature Reviews: a Guide for Students of the Social and Behavioral Sciences* (Sixth Edit). Routledge:Taylor & Francis Group. <https://doi.org/10.1016/j.wneu.2016.11.058>
- Ginting, R. S. B., & Fithriani, R. (2022). Peer and automatic writing evaluation (AWE): Indonesian EFL college students preference for essay evaluation. *LLT Journal: A Journal on Language and Language Learning*, 25(2), 461–473.
- Gutierrez, X. (2008). What Does Metalinguistic Activity in Learners' Interaction During a Collaborative L2 Writing Task Look Like? *The Modern Language Journal*, 92(4), 519–537. <https://doi.org/https://doi.org/10.1111/j.1540-4781.2008.00785.x>
- Haddaway, N. R., Page, M. J., Pritchard, C. C., & McGuinness, L. A. (2022). *PRISMA2020: An R package and Shiny app for producing PRISMA 2020-compliant flow diagrams, with*

- interactivity for optimised digital transparency and Open Synthesis Campbell Systematic Reviews*. 18, e1230. <https://doi.org/https://doi.org/10.1002/cl2.1230>
- Hong, W. C. H. (2023). The impact of ChatGPT on foreign language teaching and learning: Opportunities in education and research. *Journal of Educational Technology and Innovation*, 5(1), 37–45.
- Ismail, S. (2019). Impact of differentiated instruction on the writing process of ESL learners. *ELF Annual Research Journal*, 21, 130–153.
- Jarrah, A., Wardat, Y., & Fidalgo, P. (2023). Using ChatGPT in academic writing is (not) a form of plagiarism : What does the literature say ? Using ChatGPT in academic writing is (not) a form of plagiarism : What does the literature say ? *Online Journal of Communication and Media Technologies*, 13(4), e202346. <https://doi.org/10.30935/ojcm/13572>
- Kilkenny, M. F., & Robinson, K. M. (2018). Data quality: “ Garbage in – garbage out .” *Health Information Management Journal*, 47(3), 103–105. <https://doi.org/10.1177/1833358318774357>
- Koltovskaia, S. (2022). Postsecondary L2 writing teachers’ use and perceptions of Grammarly as a complement to their feedback. *ReCALL*, 1–15. <https://doi.org/10.1017/S0958344022000179>
- Lavallée, M., Robillard, P., & Mirsalari, R. (2014). Novices : An Iterative Approach. *IEEE Transactions on Education*, 57(3), 175–181.
- Li, Z. (2021). Teachers in automated writing evaluation (AWE) system-supported ESL writing classes: Perception, implementation, and influence. *System*, 99, 102505. <https://doi.org/10.1016/j.system.2021.102505>
- Li, Z., Dursun, A., & Hegelheimer, V. (2017). Technology and L2 Writing. In C. A. Chapelle & S. Sauro (Eds.), *The Handbook of Technology and Second Language Teaching and Learning* (1st ed., pp. 77–92). John Wiley & Sons. <https://doi.org/https://doi.org/10.1002/9781118914069.ch6>
- Liang, J. (2007). Language Scaffolding in Second Language Writing. *The Catesol Journal*, 19(1), 71–88.
- Muna, A., Ausat, A., Massang, B., Efendi, M., & Riady, Y. (2023). Can Chat GPT Replace the Role of the Teacher in the Classroom: A Fundamental Analysis. *Journal on Education*, 5(4), 16100–16106.
- Nazari, N., Shabbir, M. S., & Setiawan, R. (2021). Heliyon Application of Arti fi cial Intelligence powered digital writing assistant in higher education : randomized controlled trial. *Heliyon*, 7, e07014. <https://doi.org/10.1016/j.heliyon.2021.e07014>
- Petticrew, M., & Roberts, H. (2006). *Systematic Reviews in the Social Sciences*. Wiley-Blackwell Publishing.

- Shaffril, H. A. M., Samsuddin, S. F., & Samah, A. A. (2021). The ABC of systematic literature review : the basic methodological guidance for beginners. *Quality & Quantity*, 55, 1319–1346. <https://doi.org/10.1007/s11135-020-01059-6>
- Wee, B. Van, & Banister, D. (2015). How to Write a Literature Review Paper? *Transport Reviews*, 36(2), 278–288. <https://doi.org/10.1080/01441647.2015.1065456>
- Xiao, Y., & Watson, M. (2019). Guidance on conducting a systematic literature review. *Journal of Planning Education and Research*, 39(1), 93–112. <https://doi.org/10.1177/0739456X17723971>
- Xu, J., & Zhang, S. (2021). Understanding AWE Feedback and English Writing of Learners with Different Proficiency Levels in an EFL Classroom: A Sociocultural Perspective. *Asia-Pacific Education Researcher*, 31, 357–367. <https://doi.org/10.1007/s40299-021-00577-7>
- Yan, D. (2023). Impact of ChatGPT on learners in a L2 writing practicum: An exploratory investigation. *Education and Information Technologies*, 1–25. <https://doi.org/https://doi.org/10.1007/s10639-023-11742-4>
- Zhang, X. (2022). Improving Pigai as an Automated Writing Evaluation System: Considerations for Refinement. *Frontiers in Psychology*, 13, 795725. <https://doi.org/10.3389/fpsyg.2022.795725>

ABOUT THE AUTHORS

	Author's Biography			
	Name	<i>Angela Andrea Perez Roa</i>		
	Affiliation	<i>Universidad Nacional Abierta y a Distancia (UNAD)</i>		
	Email	<i>angela.perez@unad.edu.co</i>	Scopus ID	
	Google Scholar	<i>If any</i>	WoS ID	<i>If any</i>
	Research Interest	<i>Angela Andrea Pérez Roa is a teacher of foreign languages with an emphasis on English and French. She is a researcher for the SIBIT research group at UNAD and currently works as a teacher at a public school in Bucaramanga. Additionally, she coordinates the English area of Estudios TEC at UNAB and also teaches at UNAD.</i>		

	Author's Biography			
	Name	<i>Shanty Halim</i>		
	Affiliation	<i>Politeknik Negeri Ujung Pandang</i>		
	Email	<i>shantynurul@poliupg.ac.id</i>	Scopus ID	-
	Google Scholar	<i>Click here</i>	WoS ID	-
	Research Interest	<i>Shanty Halim is a senior lecturer of English and currently a head of the Language Center of Politeknik Negeri Ujung Pandang. Her research mainly focuses on the area of translation, ICT in ELT, AI in ELT, and applied linguistics.</i>		