

REVIEW

A Systematic Review of Automated Writing Evaluation (AWE) Systems on University Students' English Writing Performance (2020 – 2024)

Jiapeng Du ^{1,2*} , Nur Rasyidah binti Mohd Nordin ¹ 

¹ School of Languages, Civilisation and Philosophy, Universiti Utara Malaysia, Sintok 06010, Malaysia

² School of Foreign Languages, Taishan University, Tai'an 271000, China

ABSTRACT

Automated Writing Evaluation (AWE) systems, supported by advances in artificial intelligence and natural language processing, have become increasingly prominent in higher education as tools to enhance English writing instruction. This paper presents a systematic literature review of empirical studies published between 2020 and 2024 that examine the impact of AWE on university students' English writing. Following the PRISMA framework, 20 peer-reviewed studies were selected and analyzed across multiple dimensions, including writing performance, feedback types, and students' perceptions. Findings indicate that AWE systems generally improve surface-level writing features, particularly accuracy and fluency, while their effect on higher-order skills, such as content development, organization, and critical thinking remains limited. Comparative studies suggest that AWE feedback is efficient and consistent, yet often less effective than teacher feedback in fostering deep learning. Students' perceptions of AWE are mixed: many value the immediacy and convenience of feedback, while concerns about overreliance, limited adaptability, and occasional inaccuracies persist. Moreover, the lack of long-term, cross-cultural research highlights methodological and contextual gaps in the current literature. This review not only synthesizes the main contributions of existing studies but also identifies key challenges, emphasizing the need for more integrated, longitudinal, and learner-centered approaches. Several directions for future research are proposed, aiming

*CORRESPONDING AUTHOR:

Jiapeng Du, School of Languages, Civilisation and Philosophy, Universiti Utara Malaysia, Sintok 06010, Malaysia; School of Foreign Languages, Taishan University, Tai'an 271000, China; Email: jependoo@163.com

ARTICLE INFO

Received: 22 August 2025 | Revised: 10 October 2025 | Accepted: 11 October 2025 | Published Online: 21 October 2025
DOI: <https://doi.org/10.30564/fls.v7i11.11764>

CITATION

Du, J., Nordin, N.R.M., 2025. A Systematic Review of Automated Writing Evaluation (AWE) Systems on University Students' English Writing Performance (2020 – 2024). *Forum for Linguistic Studies*. 7(11): 615 – 632. DOI: <https://doi.org/10.30564/fls.v7i11.11764>

COPYRIGHT

Copyright © 2025 by the author(s). Published by Bilingual Publishing Group. This is an open access article under the Creative Commons Attribution-NonCommercial 4.0 International (CC BY-NC 4.0) License (<https://creativecommons.org/licenses/by-nc/4.0/>).

to optimize the pedagogical potential of AWE systems and support their effective integration into university-level English writing instruction.

Keywords: Automated Writing Evaluation (AWE); English Writing; Writing Performance; Feedback Types Student Perceptions; Systematic Review

1. Introduction

As a core skill of language output, English writing has irreplaceable value in academic research, professional development, and personal growth^[1]. In higher education, students need to demonstrate critical thinking skills through academic writing such as essays and reports^[2]; in the workplace environment, accurate written expression directly affects the efficiency of decision-making and professional image^[3]; and at the personal level, English writing is an important vehicle for promoting cross-cultural communication and creative expression^[4]. However, in the practice of foreign language teaching, the cultivation of writing ability is faced with a double challenge: on the one hand, teachers need to simultaneously assess multi-dimensional indicators such as grammatical accuracy, vocabulary diversity, logical coherence, and so on; on the other hand, it is difficult to guarantee the timeliness and continuity of personalised feedback under the traditional teaching mode^[5].

In this context, the rise of Automated Writing Evaluation (AWE) systems brings a paradigm change to writing teaching. Since the development of Natural Language Processing (NLP)-based technology in the 1990s^[6], AWE systems have evolved from initially correcting superficial grammatical errors to becoming intelligent tools that support formative assessment^[7]. The current mainstream systems analyse text features such as grammatical structure, lexical complexity, and chapter coherence through algorithms^[8], and their application scenarios cover a wide range of domains such as academic writing, vocational training, and have been deeply integrated into learning management systems (LMSs) and online writing platforms^[9]. Empirical studies have shown that the immediate feedback provided by AWE not only effectively improves writing accuracy^[10] but also enhances learners' writing confidence through the iterative revision mechanism^[11].

Compared to traditional teaching models, the AWE system demonstrates three core advantages: first, the instant

feedback mechanism that breaks through the limitations of time and space significantly shortens the learning cycle^[12]; second, the personalised diagnosis based on the analysis of big data can accurately locate the weaknesses in writing^[13]; and third, the stress-free writing environment helps to reduce learner anxiety^[14]. Practical examples from higher education institutions show that AWE systems have both eased instructor workloads and prompted students to shift from passive acceptance to active revision through automated corrections and learning trajectory tracking^[15].

Despite the rapid development of AWE technology, several significant gaps persist in the existing research. Most findings have focused on superficial writing skills such as grammatical error correction^[16], while the impact of AWE on higher-order competencies—such as argument logic, content organization, and critical thinking—has not been systematically examined^[17]. Moreover, the current literature tends to emphasize quantitative metrics of writing outcomes, overlooking qualitative insights into dynamic processes such as students' revision strategies and motivational changes^[18]. In addition, comparative studies between AWE and traditional feedback models, as well as the design of blended instructional approaches, remain underexplored^[9]. More notably, most existing reviews adopt a narrative structure and lack a systematic and methodologically rigorous synthesis of evidence regarding AWE efficacy across diverse teaching contexts^[19].

To date, there remains a lack of a systematic review that consolidates recent empirical evidence—especially concerning AWE's role in fostering higher-order writing skills and its comparative effectiveness against conventional feedback methods. Furthermore, many existing reviews do not adhere to established methodological guidelines such as the PRISMA framework, which is critical for ensuring transparency, reproducibility, and methodological rigor in evidence synthesis.

To address these gaps, this systematic literature review aims to critically examine empirical studies on AWE systems in university-level English writing published between

2020 and 2024. Grounded in the PRISMA framework, this review synthesizes existing research to explore the impact of AWE on students' writing performance, with a focus on the following research questions:

1. How do the AWE systems influence different dimensions of university students' English writing performance?
2. What are university students' perceptions of AWE systems?
3. How do different feedback mechanisms compare in their effectiveness in improving university students' writing performance?
4. What are the major limitations of existing AWE research, and what future directions should be explored?

2. Literature Review

2.1. Automated Writing Evaluation (AWE) Systems

With the growing status of English as a global lingua franca in academic, commercial, and technological fields, the demand for English language proficiency in higher education has significantly increased, particularly in the area of writing competence. As a result, instructors commonly adopt writing tasks accompanied by feedback mechanisms to enhance students' writing performance. However, studies have shown that providing high-quality writing evaluation imposes a considerable workload on teachers and that human scoring is often affected by rater fatigue, insufficient training, and subjective bias^[16,20,21].

To address these challenges, Automated Writing Evaluation (AWE) systems have emerged as promising tools. These systems not only alleviate teachers' workload but also improve assessment consistency and promote learner autonomy^[22,23]. The prototype of AWE can be traced back to the 1960s with the development of the Project Essay Grade (PEG) system, which utilized statistical models to analyze textual features and predict writing scores. In the 1970s, AWE began to demonstrate initial effectiveness in academic assessment and was found to achieve a level of scoring consistency comparable to that of human raters^[24].

In the digital era, advancements in computing power and internet technology have significantly accelerated the

development of AWE. Modern AWE platforms such as Grammarly, MY Access!, and Criterion go beyond basic scoring by offering multi-draft feedback and personalized suggestions, supporting learners in enhancing their writing across dimensions, including vocabulary, grammar, organization, and content^[16].

Studies have confirmed the practical value of AWE systems in standardized testing, university instruction, and academic research. Their real-time feedback and accurate scoring mechanisms have greatly improved the efficiency of writing instruction while also enhancing the reliability and validity of large-scale assessments^[25]. Notably, the adoption of AWE in foreign language education continues to expand, particularly in writing instruction, where it is playing an increasingly critical role in the optimization of formative assessment frameworks^[26,27].

2.2. Writing Performance

In Second Language Acquisition (SLA), writing performance refers to the quality of students' written output when completing authentic writing tasks^[28]. It encompasses the ability to produce meaningful, well-structured, and grammatically accurate texts to fulfill specific communicative goals. Unlike writing competence, which pertains to fundamental knowledge of writing skills, writing performance emphasizes the practical application of this knowledge in real-world contexts^[21].

Writing performance is widely recognized as a key measure of language proficiency, particularly in domains such as academic writing, professional communication, and everyday interactions. It demands not only a solid grasp of grammar and vocabulary but also the development of higher-order cognitive abilities, including critical thinking, logical reasoning, information organization, and creative expression. Consequently, writing performance is shaped not only by a learner's linguistic foundation but also by their writing strategies, affective factors, and textual structuring techniques. Kellogg et al. describe writing performance as the assessment and evaluation of written texts, determining how effectively a piece of writing conveys its intended message and meets the expectations of its audience^[29]. Therefore, writing performance can be understood as the evaluation of learners' writing based on multiple criteria, with measurable data used to represent their proficiency.

Currently, there is no universally accepted standardized framework for assessing writing performance. Researchers tend to adopt different evaluation criteria depending on the specific objectives of the writing task. For instance, Kormos assessed writing performance through multiple dimensions, including fluency, content depth, coherence, and syntactic complexity^[30]. Skehan introduced the Complexity, Accuracy, and Fluency (CAF) framework, which consists of three core dimensions—complexity, accuracy, and fluency—and is widely considered a reliable metric for evaluating second language learners' writing proficiency^[31].

2.3. Synthesizing AWE Research: The Role of Systematic Reviews

To position the current study within the existing methodological landscape, it is crucial to acknowledge the role of systematic reviews in advancing the field. Several scholars have employed the PRISMA framework to investigate technology-enhanced language learning, establishing a strong methodological precedent. For instance, Farsani et al. applied the PRISMA guidelines to review L2 argumentative writing, effectively demonstrating the framework's utility in synthesizing evidence across diverse educational settings with rigor and transparency^[32]. In terms of content-focused synthesis, Zhai and Ma conducted a meta-analysis of AWE's impact on writing quality, providing a quantitative consolidation of empirical evidence on the topic, though their study was not structured as a PRISMA-guided systematic review^[9].

The current study builds directly upon these foundational efforts. It adopts the PRISMA framework, following the methodological precedent set by researchers like Farsani et al., to systematically analyze the impact of AWE on university students' writing performance. By doing so, this review not only addresses a specific content gap but also enhances methodological rigor, replicability, and transparency, thereby contributing to a more robust and cumulative understanding of AWE's role in second language writing instruction.

3. Methodology

3.1. Study Design

This systematic literature review is conducted in accordance with the PRISMA 2020 Statement, which offers

guidelines for both the execution and reporting of systematic reviews^[33]. The PRISMA framework comprises a 27-item checklist along with a four-phase flow diagram. This study adheres to the four key phases outlined in the statement—identification, screening, eligibility, and inclusion—while ensuring compliance with the checklist requirements.

3.2. Search Strategy

This research conducts a systematic literature review examining the impact of AWE systems on the English writing performance of university students. The analysis concentrates on peer-reviewed studies published from 2020 to 2024. The chosen period is critical, as it encompasses rapid advancements in AI and NLP that have fundamentally improved AWE feedback capabilities. It also aligns with the widespread digitalization of education precipitated by the COVID-19 pandemic, which led to the rapid integration of AWE tools into online and blended learning curricula.

To ensure comprehensiveness and reliability, this study conducted a systematic search across multiple authoritative databases, including Google Scholar (GS), Scopus, ERIC (Education Resources Information Center), and CNKI (China National Knowledge Infrastructure).

The search process began with the ERIC, a European database that specializes in collecting educational research literature. Given that this systematic review focuses on the impact of AWE systems on university students' English writing performance, selecting ERIC was a logical choice, as it provides access to high-quality, peer-reviewed educational studies. However, coverage of ERIC is somewhat limited in terms of the number of available articles, particularly on technology-enhanced language learning and AWE-related research.

To broaden the scope and ensure a more comprehensive dataset, two larger interdisciplinary databases were included: GS and Scopus. GS was chosen due to its broad accessibility and extensive indexing of academic papers, including journal articles, conference proceedings, and institutional reports, making it a valuable resource for capturing a wide range of empirical studies on AWE. Scopus, as one of the largest abstract and citation databases of peer-reviewed literature, was incorporated to provide access to high-impact research across various disciplines, including education, linguistics, and artificial intelligence applications in language learning.

In addition, to enhance the geographical diversity of the research and strengthen its relevance to EFL contexts, this study included the CNKI database. Given the widespread use of AWE systems in English writing instruction at Chinese universities, the inclusion of CNKI helps supplement relevant empirical studies.

By integrating these four databases—ERIC, GS, Scopus, and CNKI—this systematic review ensures a balanced, comprehensive, and representative selection of studies, reducing publication bias while capturing a global perspective on AWE’s impact on university students’ writing perfor-

mance.

3.3. Inclusion and Exclusion Criteria

After identifying databases and completing an initial search, inclusion and exclusion criteria were developed to further screen the literature and narrow down the studies. Selected literature had to fulfil all the inclusion criteria, and any literature that met any of the exclusion criteria was excluded. The detailed inclusion and exclusion criteria are presented in **Table 1**.

Table 1. Inclusion and Exclusion Criteria.

Category	Inclusion Criteria	Exclusion Criteria
Research Topic	- Examines the impact of AWE systems on university students’ English writing performance. - Focuses on at least one aspect of writing performance (accuracy, complexity, fluency, grammar, content development, organization). - Investigates students’ attitudes and experiences with AWE.	- Focuses on other forms of writing feedback (e.g., teacher feedback, peer feedback) without incorporating AWE. - Discusses AWE technology development without analyzing its effect on writing performance.
Target Population	University students (ESL/EFL, undergraduate or graduate).	- Participants are K-12 students instead of university students. - Research primarily focuses on first-language (L1) writing, rather than ESL/EFL writing.
Research Methodology	- Empirical studies using quantitative, qualitative, or mixed methods. - Reliable data sources: The study must clearly describe data collection methods and provide data analysis.	- Non-empirical research (e.g., narrative literature reviews instead of systematic reviews, purely theoretical discussions). - Extremely small sample sizes, or unclear/unreliable data analysis methods.
AWE System	- Utilizes real AWE systems (e.g., Pigai, Criterion, Grammarly, MyAccess). - Clearly describes the role or impact of AWE feedback.	- Does not use an actual AWE system, or is a purely theoretical discussion, or focus solely on AWE system scoring quality, algorithm evaluation, or technical development. - Focuses on AI-assisted writing tools (e.g., ChatGPT) rather than AWE systems.
Time Frame	Published within 2020–2024.	Published before 2020.
Language	English Language.	Not in English.
Publication Type	Peer-reviewed journal articles and conference papers.	Unreliable sources such as preprints, blog posts, unpublished dissertations that lack peer review or formal academic indexing.

3.4. The Systematic Review Process

The systematic literature review (SLR) process began with the identification of relevant

Step 1: Keyword identification.

Keywords were selected according to the research questions and relevant literature. To improve retrieval accu-

racy, synonyms and abbreviations were incorporated into the search expressions. These search strings were repeatedly tested and refined to ensure their precision and relevance. Because different databases adopt different search mechanisms, the search strategies were adjusted accordingly.

Step 2: Database searching.

Four databases were searched: Google Scholar, Sco-

pus, ERIC, and CNKI. For Google Scholar, a title-based strategy was adopted to minimize irrelevant retrieval caused by its broad full-text search: allintitle: ('Automated Writing Evaluation' OR 'AWE' OR 'Automated Essay Scoring') AND ('university students' OR 'college students' OR 'EFL learners' OR 'ESL learners') AND ('writing performance' OR 'writing quality' OR 'writing assessment').

For Scopus and ERIC, Boolean operators were applied to TITLE-ABS-KEY fields for higher precision: (TITLE-ABS-KEY ('Automated Writing Evaluation' OR 'AWE' OR 'Automated Essay Scoring' OR 'AES' OR "Automated Writing Feedback" OR "Computer-Assisted Writing Evaluation" OR "Intelligent Writing Assessment")).

AND (TITLE-ABS-KEY ('university students' OR 'college students' OR 'EFL learners' OR 'ESL learners')) AND (TITLE-ABS-KEY ('writing performance' OR 'writing quality' OR 'writing assessment' OR 'writing proficiency' OR 'writing improvement')).

For the CNKI database, which primarily indexes Chinese literature, the following Boolean search string was applied to the "Subject" field (covering title, keywords, and abstract): SU = ('自动化写作评价 ('Automated Writing Evaluation') OR 'AWE' OR '自动作文评分 ('Automated Essay Scoring')) AND ('大学生 ('university students') OR '本科生 ('undergrad')) AND ('英语写作 ('English writing') OR '写作表现 ('writing performance') OR '写作质量 ('writing quality')). This query was designed to capture the core con-

cepts of AWE systems, the target population, and the outcome of writing performance in the Chinese academic context.

Step 3: Preliminary screening.

The initial search yielded 766 records (46 from ERIC, 264 from Scopus, 422 from Google Scholar, and 34 from CNKI). After removing duplicates, 732 records remained. Titles and abstracts were screened to exclude irrelevant studies, resulting in 259 articles for full-text review.

Step 4: Full-text review and eligibility assessment.

Full-texts were obtained through university databases, CNKI, and Google Scholar. Studies were evaluated against predefined inclusion and exclusion criteria, focusing on whether they (a) examined the impact of AWE systems on university students' writing performance, (b) contained empirical data, (c) were published in peer-reviewed journals or conference proceedings, and (d) involved EFL/ESL university learners. This step resulted in 69 studies meeting the basic inclusion criteria.

Step 5: Quality appraisal and final selection.

To ensure methodological rigor, all eligible studies underwent a quality assessment examining the clarity of research design, robustness of data analysis, and strength of theoretical grounding. Following this final screening, 20 high-quality empirical studies were included in the review.

The selection process followed the PRISMA framework, and a flowchart summarizing the exclusion at each stage is presented in **Figure 1**.

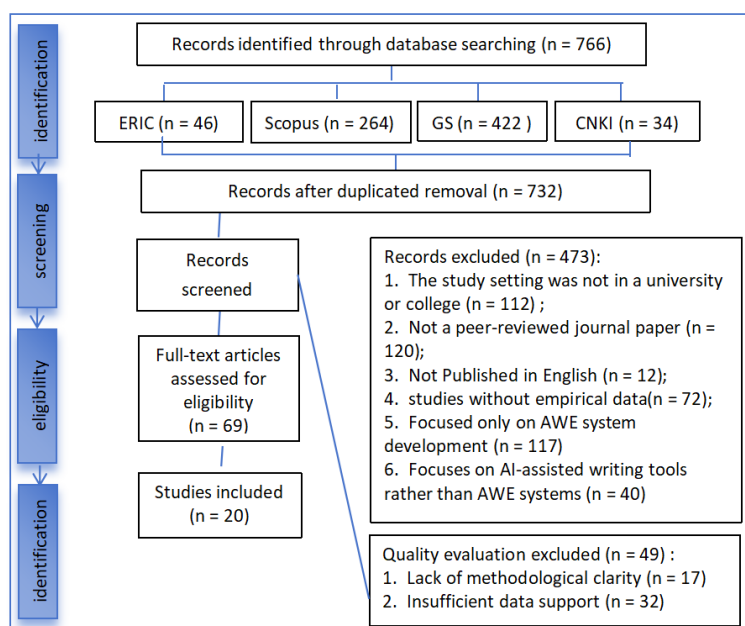


Figure 1. The PRISMA flow diagram of the literature search and review process.

3.5. Data Abstraction and Analysis

To systematically review the selected studies and address the research objective of understanding the impact of AWE systems on university students' writing performance, the screened studies were analyzed using both qualitative and quantitative methods.

A structured data extraction table was created in Excel to ensure accuracy and consistency. The extracted data encompassed the following key dimensions: (1) basic study information (e.g., authors, publication year, country); (2) participant characteristics; (3) research design and methodology; (4) AWE system utilized; (5) focus of writing performance; (6) comparison groups (if applicable); and (7) key findings related to AWE effectiveness and student perceptions. Each study was manually reviewed, and relevant information was systematically recorded into the predefined categories.

The full data extraction table (see **Appendix A**) provides a detailed overview of the included studies. The extracted data were then analyzed to identify trends, recurring themes, and gaps in the existing literature. Qualitative analysis was conducted to explore patterns in research methodologies and findings, while quantitative data were synthesized where applicable. This structured approach ensured a comprehensive understanding of how AWE influences university students' writing performance in diverse educational contexts.

3.6. Quality Assessment of Included Studies

To critically appraise the methodological rigor of the included studies, we adapted the Mixed Methods Appraisal Tool (MMAT). Each study was evaluated based on predefined criteria, including: (1) the appropriateness of the

research design for addressing the research questions, (2) the adequacy of sample size and participant recruitment, (3) the validity and reliability of data collection instruments, and (4) the thoroughness of data analysis. This assessment revealed that while most studies were well-designed, several had limitations such as small sample sizes or short intervention durations, which are acknowledged as factors that may influence the generalizability and strength of the findings discussed in this review.

4. Findings and Discussion

This section presents and discusses the results of the systematic literature review, structured around the research questions outlined in the introduction. The synthesis is based on the analysis of the 20 included studies, the full details of which are provided in **Appendix A**.

4.1. Characteristics of the Studies

A total of 20 studies were included, all published between 2020 and 2024. The distribution of publications by year is presented in **Figure 2**. All studies focused on university students' writing performance. In terms of educational level, except for two studies that focused on postgraduate students, the remaining studies all targeted undergraduate students, as shown in **Figure 3**.

In terms of geographical distribution, the studies were mainly concentrated in Asia. Only three studies were conducted outside of Asia, specifically in Turkey (2) and Egypt (1), as shown in **Figure 4**. This distribution aligns with the linguistic backgrounds of the learners involved, as illustrated in **Figure 5**.

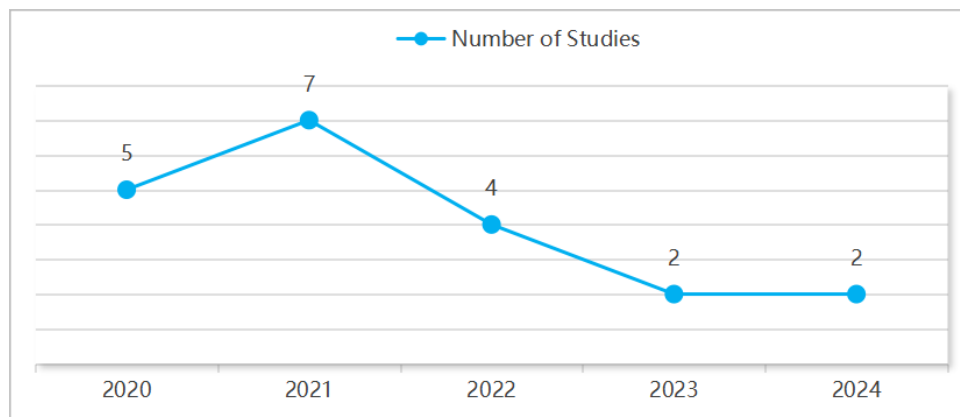


Figure 2. Publication Year Distribution.

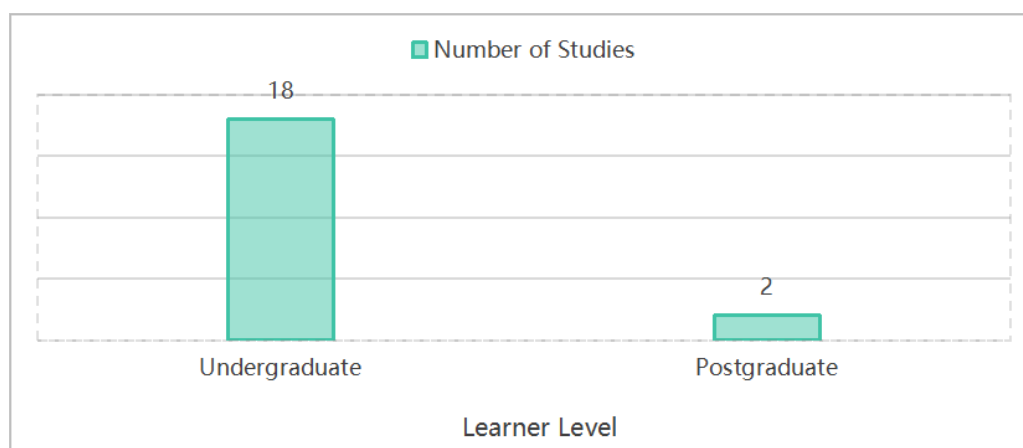


Figure 3. Learner Level Distribution.

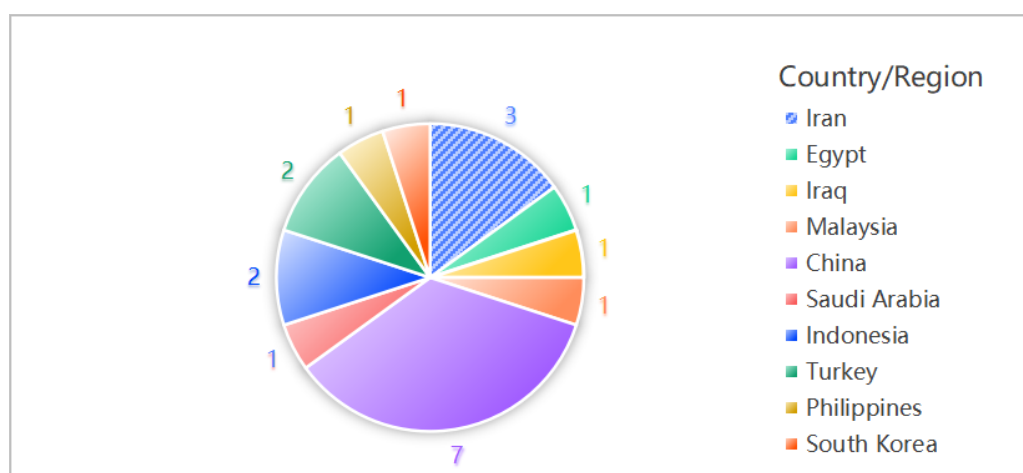


Figure 4. Geographical Distribution.

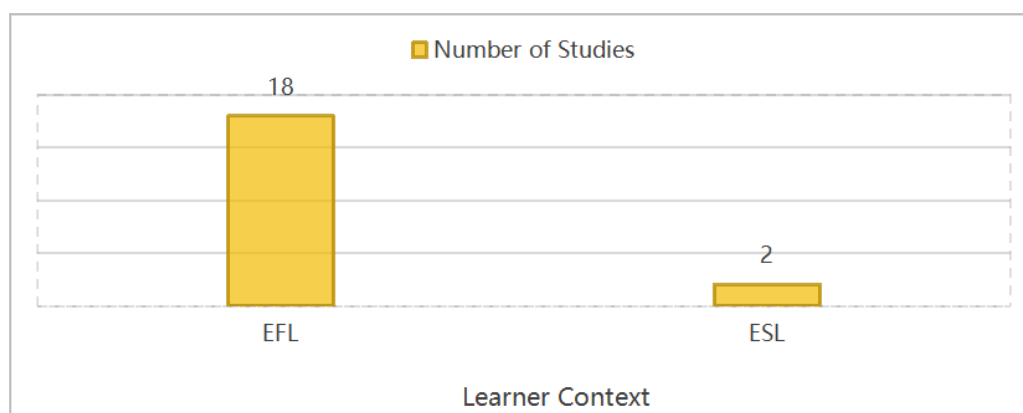


Figure 5. Learner Context Distribution.

The predominance of research in Asia may be attributed to the fact that many Asian countries are non-English-speaking regions where English is a compulsory foreign language for both academic and international communication purposes, resulting in a higher demand for AWE tools [34,35].

Additionally, due to the significant linguistic differences between Asian languages (e.g., Chinese, Arabic) and English, learners often produce higher rates of grammatical and syntactic errors, making the error correction function of AWE tools particularly valuable [36].

Among the included studies, China contributed the most (7 studies), which may be attributed to its large population, teacher-student ratio imbalance, and growing reliance on AWE systems for writing assessment. Additionally, national initiatives promoting smart education and AI integration (e.g., the 14th Five-Year Plan), along with the widespread use of localized tools like Pigai, have further driven research in this area.

The duration of the included studies ranged from four weeks to one year. Although not all studies provided specific information on the intervention period (3 studies), the major-

ity lasted more than eight weeks (13 studies), with 16-week (6 studies) duration being the most common, accounting for 69% of studies exceeding eight weeks, as shown in **Figure 6**. A 16-week duration aligns with the typical length of an academic semester, making it easier to integrate AWE into formal curricula. Moreover, short-term interventions may yield less significant outcomes compared to long-term ones, leading researchers to favor extended periods^[37]. Additionally, students require time to adapt to the feedback mechanisms of AWE tools and to internalize corrective strategies as part of their autonomous writing skills^[10].

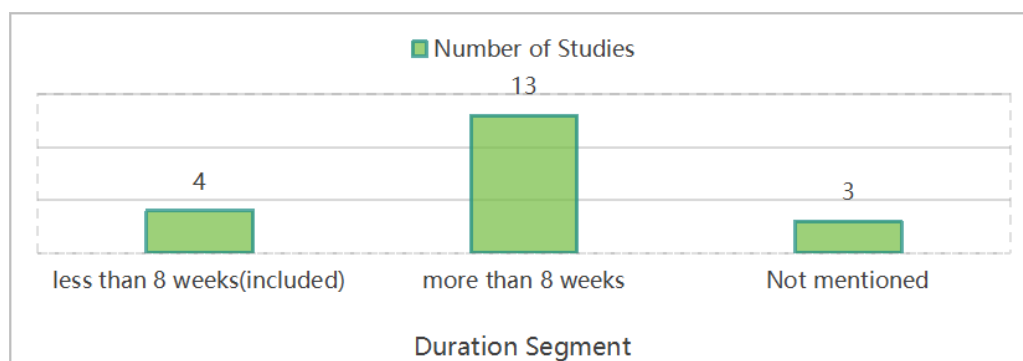


Figure 6. Study Duration Distribution.

The reviewed studies involved six different AWE systems(see **Figure 7**), with *Grammarly* (5 studies) and *Criterion* (6 studies) being the most commonly used, while studies conducted in China primarily adopted the local AWE tool *Pigai* (6 studies). *Criterion* is based on ETS’s “e-rater” engine, which also underpins the writing assessment systems of standardized tests such as TOEFL and GRE, thereby ensuring high research validity^[24]. *Grammarly*, as the most widely used AWE tool globally, is notable for

its browser-based plug-in format, which makes it the only AWE system capable of supporting informal writing contexts^[38]. *Pigai*, on the other hand, is tailored specifically to the Chinese educational context, with its scoring rubric directly aligned with the English writing assessment criteria of China’s Gaokao and CET-4/CET-6 exams^[12]. Its “templated feedback” mechanism—such as suggestions for a three-paragraph structure—closely matches exam-oriented writing requirements.

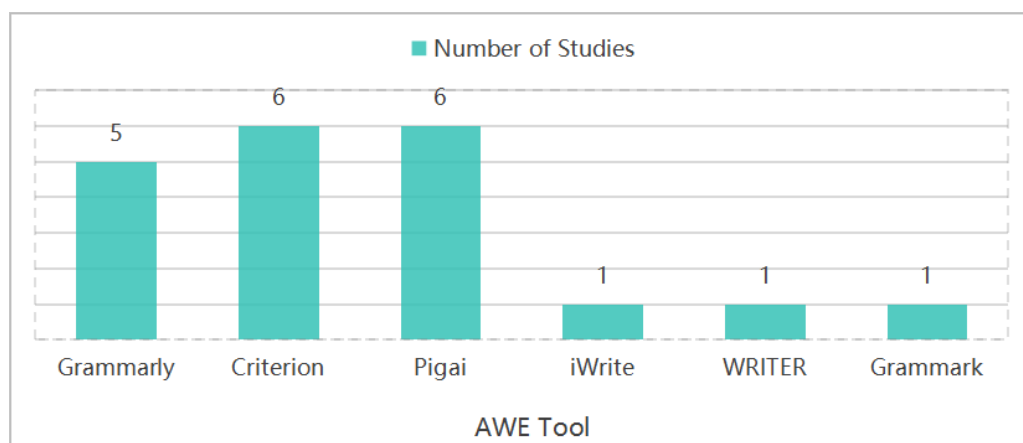


Figure 7. AWE Tools Usage Distribution.

4.2. Quantitative Effects of AWE on Writing Performance

This section focuses on the quantitative impact of AWE systems on university students' writing performance, including dimensions such as content development, organization, coherence, accuracy, complexity, and fluency. Students' perceptions, attitudes, and emotional responses toward AWE systems will be discussed under the next research question; therefore, findings related to student perceptions are not included in this section.

Through systematic literature retrieval, a total of 18 empirical studies that met the inclusion criteria were selected for this section. Among these, five were purely quantitative studies^[39–42], and thirteen were mixed-methods studies.

From a technical standpoint, research indicates that AWE systems can significantly enhance surface-level linguistic accuracy, particularly in identifying and correcting grammatical errors. Barrot's study demonstrates that systems such as Grammarly can substantially reduce severe grammatical errors but exhibit limited effectiveness in addressing minor errors (Level 1)^[43]. Empirical data from Saricaoglu and Bilki reveal that the Criterion system can reduce grammatical errors—such as subject-verb agreement issues—by 75%^[13].

Notably, while AWE improves text readability^[41], report an average increase of 1.71 points on the Flesch Readability Index), it may also lead to a decline in linguistic complexity, as evidenced by a 0.48-point decrease in grade-level scores. Link et al.'s comparative study further highlights that this simplification effect becomes more pronounced with prolonged use, underscoring the need for careful evaluation of AWE's long-term impact on linguistic sophistication^[8].

In the dimension of language structure and complexity, AWE systems show differential effects. Sun and fan confirmed that AWE has a positive impact on text organization and logical coherence^[44], but the data from Algburi et al. indicate that the improvement is relatively limited (with organization structure scores only rising from 13.11 to 16.26)^[42]. In terms of syntactic complexity, a longitudinal study by Li and Lu found that after one semester of AWE usage, T-unit length from students' work increased by 52.6%, and the use of subordinate clauses rose by 71.9%, suggesting that AWE can effectively promote the development of basic language skills^[45]. However, Chen and Cui pointed out that

this improvement mainly focuses on surface-level language features, with weaker effects on deeper discourse cohesion skills^[17]. More concerningly, Zulkornain et al. found that AWE might even lead to the simplification of sentence structures, a phenomenon that educators should be particularly cautious about^[41].

In terms of content quality, the role of AWE systems is relatively limited. Xu and Zhang found that while low-level learners can significantly improve content expression with AWE assistance^[27], Sun discovered that in terms of overall content depth and idea development, the effect of AWE is far less significant compared to teacher feedback^[44]. Data from Tian and Zhou show that students' adoption rate of teacher feedback is as high as 85.3%, while the adoption rate of automated feedback is only 24.2%^[46]. This significant gap indicates that, in fostering higher-order writing skills, AWE is currently unable to replace the personalized feedback provided by teachers. However, Diyyab's study demonstrated that when AWE is combined with the 7E teaching model, its effectiveness in improving the depth of ideas and the quality of arguments is significantly enhanced^[47].

4.3. Learners' Attitudes and Perceptions Towards AWE Tools

The data in this section are derived from two qualitative studies^[48,49] and the qualitative data from 13 mixed-methods studies. Students' perceptions of Automated Writing Evaluation (AWE) systems exhibit multidimensional characteristics, which can be summarized into the following key aspects:

Firstly, at the functional cognitive level, students' evaluations of AWE systems show clear divergence. Nazari et al. found that the vast majority of students (86.6%–97%) highly valued the practical utility of the system in grammar checking and spelling correction^[39]. Saricaoglu, Bilki, and Wan further confirmed that the real-time feedback mechanism can significantly reduce writing errors (by up to 75%)^[13,49]. However, Chen and Cui found that only 6.5% of students believed the system provided effective support for higher-order writing skills, such as content logic and coherence, which aligns with the findings of Sun, Al-Inbari, and Al-Wasy^[17,44,50]. This reflects a clear deficiency in the current AWE systems' ability to support deep-level writing.

Secondly, in terms of emotional experience, the impact of AWE systems exhibits a dual nature. Nazari et al.

found that the positive aspects include a significant increase in self-efficacy (Cohen's $d = 0.54$) and a reduction in writing anxiety (negative emotions $d = -0.98$), primarily due to the system's real-time feedback feature^[39]. This finding is also supported by Diyyab, Sari, and Han^[47,51]. However, Barrot pointed out that approximately 23% of students reported feelings of frustration caused by false positives generated by the system^[43]. Additionally, Miranti and Widiati found that upper-year students generally expressed concerns about over-reliance on the system, a conflicting sentiment that was also confirmed in the studies of Lee, Tian, and Zhou^[20,46,52]

From the psychological and behavioral dimensions, the impact of AWE systems presents multilayered complexity. On the psychological level, Sari and Han confirmed that AWE significantly enhances writing self-efficacy, particularly for anxious learners^[51]. However, Barrot's quantitative data show no significant effect on overall anxiety levels, despite students' subjective feedback indicating reduced stress^[43]. This tension between behavior and cognition is also reflected in usage patterns: students exhibit a clear "tiered correction" phenomenon—adoption rates for grammar-related corrections (99.5%) are much higher than for content-related suggestions (58.4%)^[43,49], with low-level learners being more reliant on basic correction functions^[27,48]. This tiered characteristic is related to feedback type preferences: Link et al. found that the adoption rate for teacher feedback (76%) was significantly higher than for automated feedback^[8], but AWE-assisted peer assessment was equally effective as purely manual feedback^[50]. Notably, long-term users develop systematic revision strategies, such as increasing the number of draft submissions^[20], and students consistently adopt feedback on meaning (74.8%) at a higher rate than surface-level grammar corrections (29.9%)^[46]. This provides important guidance for AWE's feedback design—there is a need to balance technical corrections with cognitive guidance.

It is noteworthy that when the AWE system is combined with collaborative learning, students' acceptance and effectiveness of use significantly improve. Wan found that 90.8% of students recognized the positive experience of using the system in a group collaboration setting^[49]. Additionally, Ginting and Fithriani^[48] demonstrated that the combination of peer assessment and AWE is the most effective feedback approach, a conclusion also supported by the comparative

study of Chen and Cui^[17].

Finally, students have expressed clear expectations for system improvements, focusing primarily on three aspects: the specification of feedback quality^[17,53], modernization of the interface design^[20,51], and the enhancement of multilingual support^[43,52]. These suggestions provide important directions for the future development of AWE systems.

Overall, the current perception of AWE systems among university students can be characterized as "technological acceptance but skepticism regarding depth"^[13,39,49]. This cognitive trait leads them to adopt an "instrumental usage" pattern^[27,44], where the system is seen as a tool for specific stages of the writing process, while higher-order writing tasks are left to human intervention^[46,50]. This finding is highly consistent with the main conclusions of 15 studies, providing a comprehensive perspective on students' complex attitudes toward AWE systems.

4.4. Comparative Studies of Different Feedback Types

In recent years, numerous quasi-experimental studies have compared the effectiveness of AWE systems with other feedback methods in EFL writing instruction. These studies generally use a pre-test-intervention-post-test quasi-experimental design, dividing students into experimental groups (using AWE) and control groups (receiving teacher or peer feedback), and evaluating the intervention effects through quantitative analysis. Among the 20 studies reviewed, 11 explored the comparison of different feedback types on university students' writing performance. Specifically, 5 studies focused on comparing AWE with teacher feedback, 2 studies investigated the effects of AWE combined with teacher feedback versus teacher feedback alone, 3 studies compared AWE with peer feedback, and 1 study examined the long-term effects of AWE, teacher feedback, and peer feedback on students' writing process. Please refer to **Table 2** for more details.

In the comparative studies between AWE and teacher feedback, the results indicate that AWE has unique advantages in improving language accuracy and revision efficiency. For example, Algburi et al. found that the Criterion system, when integrated with the process approach to writing, significantly improved the writing quality of Iraqi undergraduates,

even outperforming traditional teacher feedback (effect size $\eta^2 = 0.86$)^[42]. Additionally, Barrot showed that Grammarly effectively reduces serious grammatical errors but has limited improvement on minor mistakes like spelling errors. Furthermore, real-time feedback from AWE and standardized evaluation not only enhances revision efficiency but also

promotes students' autonomous learning through metalinguistic explanations^[43]. This advantage has been validated in both business English writing and academic writing^[42,44]. However, despite AWE's strong performance in language accuracy, its support for content development and text organization remains limited.

Table 2. Distribution of Studies by Feedback Type.

Feedback Type	Number of Studies	Researchers
AWE vs. Teacher Feedback	5	Algburi et al., 2024 ^[42] ; Zulkornain et al., 2023 ^[41] ; Hassanzadeh & Fotoohnejad, 2021 ^[40] ; Barrot, 2021 ^[43] ; Sun & Fan, 2022 ^[44] .
AWE + Teacher Feedback vs. Teacher Feedback	2	Link et al., 2020 ^[8] ; Sari & Han, 2024 ^[51] .
AWE vs. Peer Feedback	3	Chen & Cui, 2022 ^[17] ; Xie et al., 2020 ^[53] ; Ginting & Fithriani, 2022 ^[48] .
AWE vs. Teacher vs. Peer Feedback	1	Tian & Zhou, 2020 ^[46] .

In addition to the comparative studies between AWE and teacher feedback, two studies explored the differences between AWE combined with teacher feedback and teacher feedback alone. Link et al. focused on feedback behavior and revision practices, finding that although the use of AWE reduced the amount of low-level feedback from teachers, it did not increase high-level feedback, and students tended to ignore AWE feedback at a relatively high rate (around 24%)^[8]. In contrast, Sari and Han focused on psychological factors and found that AWE combined with teacher feedback significantly improved students' writing self-efficacy and performance but had no significant impact on writing anxiety. In terms of writing performance, both studies found that AWE combined with teacher feedback was superior to teacher feedback alone in terms of language accuracy^[51], while Link et al. further pointed out that teacher feedback alone had an advantage in language complexity^[8]. This suggests that AWE can serve as a supplementary tool to teacher feedback, especially for correcting lower-level language errors, but the quality of feedback still needs to be optimized, and efforts should be made to help students better understand and apply the feedback.

In the comparative studies between AWE and peer feedback, existing research generally recognizes that both have their respective advantages. Chen and Cui found that although the iWrite system provides real-time feedback, it has limited effectiveness in improving discourse coherence and cohesion, with students generally perceiving its feedback as too vague^[17]. Peer feedback, on the other hand, was more

effective in establishing cohesion chains and improving text coherence^[16]. Xie et al. further found that the Pigai system significantly enhanced lexical diversity and syntactic complexity, but had little impact on discourse coherence, while peer feedback, though improving writing scores and accuracy, had a more limited role in language complexity^[53]. Additionally, Ginting and Fithriani conducted a qualitative case study and found that while Indonesian EFL students acknowledged the practicality of Grammarly for quick grammar checks, they preferred peer feedback, believing it offered more interactive and targeted suggestions^[48]. These findings suggest that AWE is more suitable for language form corrections, whereas peer feedback holds greater potential in developing higher-order writing skills.

In addition to the aforementioned comparative studies, Tian and Zhou conducted a 17-week case study in a naturalistic setting, exploring the use of AWE, peer feedback, and teacher feedback among Chinese university students with varying levels of English proficiency in an online EFL writing course. The study found that teacher feedback had the highest adoption rate (85.3%), especially in terms of content revisions. Peer feedback had a moderate adoption rate (55.3%), with its value influenced by peer relationships, while AWE feedback had the lowest adoption rate (24.2%), primarily used for correcting grammatical errors^[46]. The study also indicated that learners' feedback utilization patterns were dynamic, influenced by behavioral, emotional, and cognitive factors.

4.5. Theoretical Interpretation of Findings

The integrated perspectives of Formative Assessment Theory, Social Constructivism, Learner Autonomy, and the Control-Value Theory of Achievement Emotions together provide a coherent interpretive framework for the findings presented in Sections 4.2–4.4.

The quantitative evidence demonstrates that AWE systems are most effective in enhancing surface-level linguistic accuracy, which aligns with their role as formative assessment tools that provide immediate, criterion-referenced feedback to guide improvement. However, their limited contribution to content development and discourse organization reveals a lack of support for higher-order self-regulation—a core mechanism of formative learning^[54]. This imbalance indicates that AWE fosters procedural accuracy but falls short in promoting metacognitive engagement and reflective meaning construction.

Students' strategic use of AWE systems further illustrates the principles of learner autonomy: they employ automated feedback for low-level corrections while deliberately relying on human feedback for complex cognitive tasks. Such selective engagement suggests that learners are not passive recipients of feedback but active regulators of their own learning processes.

The comparative advantage of teacher and peer feedback in improving coherence and idea development can be interpreted through Social Constructivism, which emphasizes dialogic meaning-making and scaffolding within the learner's Zone of Proximal Development (ZPD)^[55]. When AWE feedback is embedded within this collaborative ecology, its efficiency and objectivity complement the interpretive depth of human mediation, leading to more holistic writing development.

Finally, the Control-Value Theory provides insight into the emotional ambivalence observed in learners' responses. While the immediacy of AWE feedback enhances perceived control and strengthens self-efficacy, its impersonal and mechanical nature can sometimes diminish perceived task value and trigger frustration.

4.6. Limitations of Current Research and Future Directions

Despite the growing body of research on AWE systems, several notable limitations remain. First, the geographical

distribution of studies is relatively narrow, with a strong focus on Asian EFL contexts, particularly in China^[9,19], which limits the generalizability of findings to broader ESL or multilingual populations. Moreover, most participants are English majors, while learners from other academic disciplines or sociocultural contexts remain underrepresented^[56].

Second, current research primarily emphasizes lower-order writing features such as grammatical accuracy and lexical diversity^[40], while higher-order skills like critical thinking, logical structuring, and academic expression are often overlooked^[19]. Methodological issues also persist, including small sample sizes, short intervention durations, and lack of control groups, which weaken the internal validity of findings^[53,56].

Additionally, few studies have tracked students' actual uptake of AWE feedback or their cognitive engagement during revision, leaving the learning mechanisms behind feedback adoption underexplored^[9]. While affective variables such as writing anxiety and self-efficacy have been studied^[57], results remain inconsistent and sometimes contradictory.

Future research should broaden participant diversity, explore the development of higher-order writing abilities, and apply multidimensional evaluation frameworks such as the CAF model^[19]. The use of mixed methods and longitudinal designs is also recommended to assess the long-term effects of AWE. Furthermore, studies should investigate students' feedback literacy and how teachers integrate AWE systems into instructional practices^[57].

5. Conclusions

This study conducted a systematic review of 20 peer-reviewed empirical studies published between 2020 and 2024, following the PRISMA framework, to examine the impact of AWE systems on university students' English writing performance. The review focused on key dimensions such as linguistic accuracy, syntactic complexity, feedback uptake, and student perceptions, and compared AWE with teacher and peer feedback.

Findings indicate that AWE systems are particularly effective in enhancing grammatical accuracy and syntactic complexity. Their real-time feedback and personalized diagnostic features help improve writing efficiency and learner confidence, especially for anxiety-prone students. AWE also alleviates teacher workload and offers scalable solutions for writing assessment.

However, limitations remain. AWE feedback primarily targets surface-level features, with limited support for content development and logical coherence. Adoption rates are sometimes low due to vague or mechanical feedback. Cultural and individual differences also affect students' trust in and use of AWE systems. Moreover, most existing studies are concentrated in Asian EFL contexts, focus on English majors, and rely heavily on quantitative methods, lacking longitudinal and qualitative depth.

Future research should diversify participant backgrounds, including non-English majors and learners from varied linguistic and cultural contexts. More attention is needed on the potential of AWE to foster higher-order writing skills such as argumentation and rhetorical organization. Mixed-methods and longitudinal designs are recommended to better understand students' engagement with automated feedback. In addition, future work should explore strategies for integrating AWE into classroom instruction and enhancing students' feedback literacy, promoting more effective and equitable writing pedagogy.

Author Contributions

Conceptualization, methodology, data collection and extraction, formal analysis, writing—original draft prepara-

tion, and editing, J.D.; supervision, resources, manuscript review and revision, and project administration, N.R.b.M.N. All authors have read and agreed to the published version of the manuscript.

Funding

This work received no external funding.

Institutional Review Board Statement

Not applicable.

Informed Consent Statement

Not applicable.

Data Availability Statement

The data used in this study are available from the corresponding author upon reasonable request.

Conflicts of Interest

The authors declare no conflict of interest.

Appendix A

Table A1. Research Matrix for AWE studies.

Author(s) & Year	Country	Participants	AWE Tool	Comparison Group	Research Method	Duration	Focus	Key Findings
Nazari., (2021)	Iran	120 EFL postgraduates	Grammarly	AWE-group vs. Non-AWE group	Quantitative method	12 weeks	Engagement, self-efficacy, emotions	Improved confidence and autonomy; reduced cognitive load
Algburi et al, (2024)	Iraq	52 EFL undergraduates	Criterion	AWE feedback vs. Teacher feedback	Quantitative method	10 weeks	Content development, organization	AWE group outperformed in content development
Chen & Cui (2022)	China	96 EFL undergraduates	iWrite	AWE feedback vs. Peer feedback	Mixed method	4 weeks	Coherence and cohesion	iWrite was unclear and ineffective in improving cohesion
Xie et al., (2020)	China	64 EFL non-English majors	Pigai	AWE feedback vs. Peer feedback	Mixed method	16 weeks	Accuracy, lexical diversity, syntactic complexity	AWE more effective in vocabulary and complexity

Table A1. *Cont.*

Author(s) & Year	Country	Participants	AWE Tool	Comparison Group	Research Method	Duration	Focus	Key Findings
Delsa Miranty & Utami Widiati (2021)	Indonesia	100 EFL undergraduates	Grammarly	None	Mixed method	—	Students' perceptions by year level	Frequency influenced acceptance;
Wenji Li & Zhihong Lu (2020)	China	66 EFL freshmen non-English majors	Pigai	None	Mixed method	5 weeks	Syntactic complexity	Increased sentence length, subordination, coordination
Jinfen Xu & Shanshan Zhang (2021)	China	65 EFL sophomores	Pigai	None	Mixed method	15 weeks	Accuracy, complexity, fluency	Improved accuracy, no effect on complexity
Rita Seroja Br Ginting & Rahmah Fithriani (2022)	Indonesia	20 EFL undergraduates	Grammarly	AWE feedback vs. Peer feedback	Qualitative method	16 weeks	Feedback frequency and preference	Peer feedback valued for its authenticity
Mohammad Hassanzadeh & Samira Fotoohnejad (2021)	Iran	53 EFL undergraduates	Criterion	AWE feedback vs. Teacher feedback	Quantitative method	12 weeks	Writing quality	AWE more consistent and less burdensome than teachers
Lili Tian & Yu Zhou (2020)	China	5 EFL sophomores	Pigai	AWE vs. Peer vs. Teacher feedback	Mixed method	17 weeks	Feedback engagement	AWE had highest feedback quantity but lowest uptake (29.9%)
Eman Aly El-ssayed Diyyab (2021)	Egypt	30 EFL second-year education majors	Grammarly	None	Mixed method	—	Academic writing skills	Improved structure, organization, language use
Xinxin Wan (2024)	China	250 EFL university students	Pigai	None	Qualitative method	18 weeks	Attitudes towards AWE tool	Positive attitudes influenced by tools
Luqmanul Hakim Zulkornain et al. (2023)	Malaysia	121 ESL undergraduates	Grammarly	AWE feedback vs. Teacher feedback	Quantitative method	10 weeks	Writing skills and readability	Skill improved, readability improvement limited
Al-Inbari, F. A. Y., & Al-Wasy, B. Q. M. (2023)	Saudi Arabia	44 EFL students	WRITER	AWE-aided vs. unaided revision	Mixed method	—	Peer/self-revision outcomes	AWE positively supported both revision types
Bo Sun & Tingting Fan (2022)	China	73 EFL business undergraduates	Pigai	AWE feedback vs. Teacher feedback	Mixed method	16 weeks	Writing performance, anxiety	Improved organization; anxiety unchanged
Link, S., Mehrzad, M., & Rahimi, M. (2020)	Iran	28 EFL students	Criterion	AWE + teacher feedback vs. Teacher feedback	Quantitative method	16 weeks	Revision behavior, writing improvement	Improved accuracy; complexity unchanged
Elif Sari & Turgay Han (2023)	Turkey	79 EFL undergraduates	Criterion	AWE + teacher feedback vs. Teacher feedback	Mixed method	16 weeks	Self-efficacy, regulation, anxiety	Boosted self-efficacy; anxiety unchanged
Jessie S. Barrot (2021)	Philippines	70 ESL students	Grammarly	AWE feedback vs. Teacher feedback	Mixed method	14 weeks	Accuracy	Reduced spelling and punctuation errors

Table A1. *Cont.*

Author(s) & Year	Country	Participants	AWE Tool	Comparison Group	Research Method	Duration	Focus	Key Findings
Lee, YJ. (2020)	South Korea	2 EFL under-graduates	Criterion	None	Mixed method	1 year	TOEIC writing score, fluency	Enhanced grammar awareness and revision skills
Saricaoglu, A. & Bilki Z. (2021)	Turkey	114 EFL under-graduates	Criterion	None	Mixed method	16 weeks	Grammatical accuracy	Reduced subject-verb agreement and other errors

References

- [1] Richards, J.C., 2010. Competence and performance in language teaching. *RELC Journal*. 41(2), 101 – 122. DOI: <https://doi.org/10.1177/0033688210372953>
- [2] Zhang, X., 2018. Developing college EFL writers' critical thinking skills through online resources: A case study. *SAGE Open*. 8(4), Article 2158244018820386. DOI: <https://doi.org/10.1177/2158244018820386>
- [3] Schriver, K., 2012. What we know about expertise in professional communication. In: Berninger, V.W. (Ed.). *Past, Present, and Future Contributions of Cognitive Writing Research to Cognitive Psychology*, 1st ed. Psychology Press: New York, NY, USA. pp. 275 – 312.
- [4] Salomone, R., Salomone, R.C., 2022. *The Rise of English: Global Politics and the Power of Language*. Oxford University Press: New York, NY, USA. pp. 45 – 62.
- [5] Whalley, B., France, D., Park, J., et al., 2021. Towards flexible personalized learning and the future educational system in the fourth industrial revolution in the wake of COVID-19. *Higher Education Pedagogies*. 6(1), 79 – 99. DOI: <https://doi.org/10.1080/23752696.2021.1883458>
- [6] Burstein, J., Chodorow, M., 1999. Automated essay scoring for nonnative English speakers. In *Proceedings of a Symposium on Computer Mediated Language Assessment and Evaluation in Natural Language Processing—ASSESEVALNLP'99*, College Park, MD, USA, 22 June 1999; pp. 68 – 75.
- [7] Koltovskaia, S., 2022. Automated writing evaluation for formative second language assessment: Exploring performance, teacher use, and student engagement [Doctoral dissertation]. Oklahoma State University: Stillwater, OK, US.
- [8] Link, S., Mehrzad, M., Rahimi, M., 2022. Impact of automated writing evaluation on teacher feedback, student revision, and writing improvement. *Computer Assisted Language Learning*. 35(4), 605 – 634. DOI: <https://doi.org/10.1080/09588221.2020.1743323>
- [9] Zhai, N., Ma, X., 2023. The effectiveness of automated writing evaluation on writing quality: A meta-analysis. *Journal of Educational Computing Research*. 61(4), 875 – 900. DOI: <https://doi.org/10.1177/07356331221127300>
- [10] Stevenson, M., Phakiti, A., 2019. Automated feedback and second language writing. In: Hyland, K., Hyland, F. (Eds.). *Feedback in Second Language Writing: Contexts and issues*, 2nd ed. Cambridge University Press: Cambridge, UK. pp. 125 – 142.
- [11] Li, A.W., Hebert, M., 2024. Unpacking an online peer-mediated and self-reflective revision process in second-language persuasive writing. *Reading and Writing*. 37(6), 1545 – 1573. DOI: <https://doi.org/10.1007/s11145-023-10466-8>
- [12] Li, J., Link, S., Hegelheimer, V., 2015. Rethinking the role of automated writing evaluation (AWE) feedback in ESL writing instruction. *Journal of Second Language Writing*. 27, 1 – 18. DOI: <https://doi.org/10.1016/j.jslw.2014.10.004>
- [13] Saricaoglu, A., Bilki, Z., 2021. Voluntary use of automated writing evaluation by content course students. *ReCALL*. 33(3), 265 – 277. DOI: <https://doi.org/10.1017/S0958344021000021>
- [14] Gochitashvili, K., 2023. Creating a safe and stress-free environment when using extracurricular activities in second language teaching. *International Journal of Multilingual Education*. (23), 1 – 15.
- [15] Grimes, D., Warschauer, M., 2021. Utility in a fallible tool: A multi-site case study of automated writing evaluation. *The Journal of Technology, Learning and Assessment*. 8(6). Available from: <https://ejournals.bc.edu/index.php/jtla/article/view/1625>
- [16] McNamara, D.S., Kendeou, P., 2022. The early automated writing evaluation (eAWE) framework. *Assessment in Education: Principles, Policy & Practice*. 29(2), 150 – 182. DOI: <https://doi.org/10.1080/0969594X.2022.2037509>
- [17] Chen, M., Cui, Y., 2022. The effects of AWE and peer feedback on cohesion and coherence in continuation writing. *Journal of Second Language Writing*. 57, 100915. DOI: <https://doi.org/10.1016/j.jslw.2022.100915>
- [18] Cope, B., Kalantzis, M., McCarthy, S., et al., 2011. Technology-mediated writing assessments: Principles and processes. *Computers and Composition*. 28(2), 79 – 96. DOI: <https://doi.org/10.1016/j.compcom.2011.04.007>

- [19] Zhang, Z.V., Hyland, K., 2018. Student engagement with teacher and automated feedback on L2 writing. *Assessing Writing*. 36, 90 – 102. DOI: <https://doi.org/10.1016/j.asw.2018.02.004>
- [20] Lee, Y.J., 2020. The long-term effect of automated writing evaluation feedback on writing development. *English Teaching*. 75(1), 67 – 92. DOI: <https://doi.org/10.15858/engtea.75.1.202003.67>
- [21] Weigle, S.C., 2013. *Assessing Writing*. Cambridge University Press: Cambridge, UK. pp. 58 – 77.
- [22] Shermis, M.D., Burstein, J., 2013. *Handbook of Automated Essay Evaluation*, 1st ed. Routledge: New York, NY, USA. pp. 36 – 55.
- [23] Warschauer, M., Ware, P., 2006. Automated writing evaluation: Defining the classroom research agenda. *Language Teaching Research*. 10(2), 157 – 180. DOI: <https://doi.org/10.1191/1362168806lr190oa>
- [24] Attali, Y., Burstein, J., 2006. Automated essay scoring with e-rater® V.2. *The Journal of Technology, Learning and Assessment*. 4(3). Available from: <https://ejournals.bc.edu/index.php/jtla/article/view/1650>
- [25] Messer, M., Brown, N.C., Kölling, M., et al., 2024. Automated grading and feedback tools for programming education: A systematic review. *ACM Transactions on Computing Education*. 24(1), 1 – 43. DOI: <https://doi.org/10.1145/3636515>
- [26] Fu, Q.K., Zou, D., Xie, H., et al., 2022. A review of AWE feedback: Types, learning outcomes, and implications. *Computer Assisted Language Learning*. 37(1 – 2), 179 – 221. DOI: <https://doi.org/10.1080/09588221.2022.2033787>
- [27] Xu, J., Zhang, S., 2022. Understanding AWE feedback and English writing of learners with different proficiency levels in an EFL classroom: A socio-cultural perspective. *The Asia-Pacific Education Researcher*. 31, 357 – 367. DOI: <https://doi.org/10.1007/s40299-021-00577-7>
- [28] Faigley, L., 1985. Nonacademic writing: The social perspective. In: Odell, L., Goswami, D. (Eds.). *Writing in Nonacademic Settings*. The Guilford Press: New York, NY, USA. pp. 231 – 248.
- [29] Kellogg, R.T., Whiteford, A.P., Turner, C.E., et al., 2013. Working memory in written composition: An evaluation of the 1996 model. *Journal of Writing Research*. 5(2), 159 – 190. DOI: <https://doi.org/10.17239/jowr-2013.05.02.1>
- [30] Kormos, J., 2011. Task complexity and linguistic and discourse features of narrative writing performance. *Journal of Second Language Writing*. 20(2), 148 – 161. DOI: <https://doi.org/10.1016/j.jslw.2011.02.001>
- [31] Skehan, P., 1998. *A Cognitive Approach to Language Learning*. Oxford University Press: New York, NY, USA. pp. 75 – 88.
- [32] Farsani, M.A., Stapleton, P., Jamali, H.R., 2025. Charting L2 argumentative writing: A systematic review. *Journal of Second Language Writing*. 68, 101208. DOI: <https://doi.org/10.1016/j.jslw.2025.101208>
- [33] Page, M.J., McKenzie, J.E., Bossuyt, P.M., et al., 2021. The PRISMA 2020 statement: an updated guideline for reporting systematic reviews. *The British Medical Journal*. 372, Article n71. DOI: <https://doi.org/10.1136/bmj.n71>
- [34] Chen, C.F., Cheng, W.Y., 2008. Beyond the design of automated writing evaluation: Pedagogical practices and perceived learning effectiveness. *Language Learning & Technology*. 25(2), 94 – 112.
- [35] Hu, G., McKay, S.L., 2012. English language education in East Asia: Some recent developments. *Journal of Multilingual and Multicultural Development*. 33(4), 345 – 362. DOI: <https://doi.org/10.1080/01434632.2012.661434>
- [36] Namaziandost, E., Sawalmeh, M.H.M., Soltanabadi, M.I., 2020. The effects of spaced versus massed distribution instruction on EFL learners' vocabulary recall and retention. *Cogent Education*. 7(1), Article 1792261. DOI: <https://doi.org/10.1080/2331186X.2020.1792261>
- [37] Koltovskaia, S., 2020. Student engagement with automated written corrective feedback (AWCF) provided by Grammarly: A multiple case study. *Assessing Writing*. 44, 100450. DOI: <https://doi.org/10.1016/j.asw.2020.100450>
- [38] Dizon, G., Gayed, J.M., 2024. A systematic review of Grammarly in L2 English writing contexts. *Cogent Education*, 11(1). DOI: <https://doi.org/10.1080/2331186X.2024.2397882>
- [39] 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), Article e07014. DOI: <https://doi.org/10.1016/j.heliyon.2021.e07014>
- [40] Hassanzadeh, M., Fotoohnejad, S., 2021. Implementing an automated feedback program for a foreign language writing course: A learner-centric study. *Journal of Computer Assisted Learning*. 37(5), 1494 – 1507. DOI: <https://doi.org/10.1111/jcal.12587>
- [41] Zulkornain, L.H., Mat, A.C., Rahman, N.A.A., 2023. Does automated writing evaluation (AWE) improve students' writing? Focus on technical aspects and readability. *International Journal of Information and Education Technology*. 13(10), 1656 – 1662.
- [42] Algburi, E., Razali, A.B., Nimehchisalem, V., et al., 2024. Combination of AWE (Criterion) feedback with the process approach and its impact on EFL writing content/idea development and organization. *International Journal of Academic Research in Progressive Education and Development*. 13(1), 793 – 803. DOI: <http://dx.doi.org/10.6007/IJARPE/v13-i1/20082>
- [43] Barrot, J.S., Llenares, I.I., del Rosario, L.S., 2021. Students' online learning challenges during the pan-

- demic and how they cope with them: The case of the Philippines. *Education and Information Technologies*. 26, 7321 – 7338. DOI: <https://doi.org/10.1007/s10639-021-10589-x>
- [44] Sun, B., Fan, T., 2022. The effects of an AWE-aided assessment approach on business English writing performance and writing anxiety: A contextual consideration. *Studies in Educational Evaluation*. 72, Article 101123. DOI: <https://doi.org/10.1016/j.stueduc.2021.101123>
- [45] Li, W., Lu, Z., Liu, Q., 2020. Syntactic complexity development in college students' essay writing based on AWE. In: Frederiksen, K.-M., Larsen, S., Bradley, L., et al. (Eds.). *CALL for Widening Participation: Short Papers from EUROCALL*. Research-publishing.net: Paris, France. pp. 190 – 194. DOI: <https://doi.org/10.14705/rpnet.2020.48.1187>
- [46] Tian, L., Zhou, Y., 2020. Learner engagement with automated feedback, peer feedback and teacher feedback in an online EFL writing context. *System*. 91, Article 102247. DOI: <https://doi.org/10.1016/j.system.2020.102247>
- [47] Diyyab, E.A., Aly, E., 2021. A suggested program based on 7E instructional model and AWE systems to develop faculty of education ESP students' academic writing skills. *Journal of Education — Sohag University*. (88)88, 1 – 39.
- [48] Ginting, R.S., Fithriani, R., 2022. Peer and automated writing evaluation (AWE): Indonesian EFL college students' preference for essay evaluation. *LLT Journal: A Journal on Language and Language Teaching*. 25(2), 461 – 473. DOI: <https://doi.org/10.24071/llt.v25i2.4879>
- [49] Wan, X., 2024. Learning English writing collaboratively with AWE: Chinese EFL learners' perceptions, engagement, and scaffolding [Doctoral dissertation]. University of Macau: Macau, China.
- [50] Al-Inbari, F.A.Y., Al-Wasy, B.Q.M., 2023. The impact of automated writing evaluation (AWE) on EFL learners' peer and self-editing. *Education and Information Technologies*. 28(6), 6645 – 6665. DOI: <https://doi.org/10.1007/s10639-022-11458-x>
- [51] Sari, E., Han, T., 2024. The impact of automated writing evaluation on English as a foreign language learners' writing self-efficacy, self-regulation, anxiety, and performance. *Journal of Computer Assisted Learning*. 40(5), 2065 – 2080. DOI: <https://doi.org/10.1111/jcal.13004>
- [52] Miranty, D., Widiati, U., 2021. Automated writing evaluation (AWE) in higher education: Indonesian EFL students' perceptions about Grammarly use across student cohorts. *Pegem Journal of Education and Instruction*. 11(4), 126 – 137. DOI: <https://doi.org/10.47750/pegog.11.04.12>
- [53] Xie, Y., Huang, L., Wang, Y., 2020. The impact of AWE and peer feedback on Chinese EFL learners' English writing performance. In: Lee, L.K., U, L.H., Wang, F.L., et al. (Eds.). *Technology in Education: Innovations for Online Teaching and Learning*. ICTE 2020: Communications in Computer and Information Science, vol 1302. Springer: Singapore. pp. 258 – 270. DOI: https://doi.org/10.1007/978-981-33-4594-2_22
- [54] Sadler, D.R., 1989. Formative assessment and the design of instructional systems. *Instructional science*. 18(2), 119 – 144.
- [55] Hyland, K., Hyland, F., 2006 Feedback on second language students' writing. *Language Teaching*. 39(2), 83 – 101. Available from: <https://core.ac.uk/download/pdf/37893651.pdf>
- [56] Suh, E.K., Dyer, J., McGee, B., et al., 2022. To, through, and beyond higher education: A literature review of multilingual immigrant students' community college transitions. *Community College Journal of Research and Practice*. 46(5), 301 – 317. DOI: <https://doi.org/10.1080/10668926.2020.1841045>
- [57] Sabti, A.A., Md Rashid, S., Nimehchisalem, V., et al., 2019. The Impact of Writing Anxiety, Writing Achievement Motivation, and Writing Self-Efficacy on Writing Performance: A Correlational Study of Iraqi Tertiary EFL Learners. *Sage Open*. 9(4). DOI: <https://doi.org/10.1177/2158244019894289>