



OPEN ACCESS

EDITED BY

Sonsoles López-Pernas,
University of Eastern Finland, Finland

REVIEWED BY

Ramiza Haji Darmi,
Putra Malaysia University, Malaysia
Carolina Cordeiro,
University of Porto, Portugal
Sara Kashefian-Naeeni,
Shiraz University of Medical Sciences, Iran

*CORRESPONDENCE

Ping Wei

✉ peiqi469@163.com

RECEIVED 29 June 2023

ACCEPTED 06 September 2023

PUBLISHED 29 September 2023

CITATION

Wei P, Wang X and Dong H (2023) The impact of automated writing evaluation on second language writing skills of Chinese EFL learners: a randomized controlled trial.
Front. Psychol. 14:1249991.
doi: 10.3389/fpsyg.2023.1249991

COPYRIGHT

© 2023 Wei, Wang and Dong. This is an open-access article distributed under the terms of the [Creative Commons Attribution License \(CC BY\)](https://creativecommons.org/licenses/by/4.0/). The use, distribution or reproduction in other forums is permitted, provided the original author(s) and the copyright owner(s) are credited and that the original publication in this journal is cited, in accordance with accepted academic practice. No use, distribution or reproduction is permitted which does not comply with these terms.

The impact of automated writing evaluation on second language writing skills of Chinese EFL learners: a randomized controlled trial

Ping Wei*, Xiaosai Wang and Hui Dong

School of Foreign Languages, Tangshan Normal University, Tangshan, Hebei, China

Introduction: In the context of the burgeoning field of second language (L2) education, where proficient writing plays an integral role in effective language acquisition and communication, the ever-increasing technology development has influenced the trajectory of L2 writing development.

Methods: To address the need for enhanced writing skills among English as a Foreign Language (EFL) learners, this study investigates the efficacy of Automated Writing Evaluation (AWE) training. A randomized controlled trial employing repeated measures was conducted, involving a participant pool of 190 Chinese EFL students. The study comprehensively assessed the effects of AWE training, utilizing the Grammarly platform—an AI-driven program—on various dimensions of writing skills, encompassing task achievement, coherence and cohesion, lexicon, and grammatical accuracy. Control variables included writing self-efficacy and global English proficiency. Writing skills were evaluated through the administration of an International English Language Testing System (IELTS) writing sample test.

Results: The results unequivocally demonstrate that the experimental group consistently exhibited superior performance across all facets of writing skills compared to the control group. Furthermore, the predictive influence of pre-test scores was pronounced in task achievement, coherence and cohesion, and lexicon, highlighting the pivotal role of learners' initial proficiency levels in shaping subsequent writing outcomes. Notably, the emergence of writing self-efficacy as a significant predictor of task achievement and coherence and cohesion underscores the role of learners' beliefs and confidence in shaping their writing abilities.

Discussion: These findings conclusively suggest that Artificial Intelligence-based instructional programs, specifically AWE, hold the potential to effectively enhance second language writing skills, especially among learners with lower proficiency levels. This study carries crucial implications for EFL educators and researchers, advocating for the seamless integration of AWE into pedagogical strategies to foster a marked improvement in writing competence.

KEYWORDS

automated writing evaluation, L2 writing skills, EFL learners, randomized controlled trial, writing self-efficacy, complex script

1. Introduction

The integration of technology in language learning has gained increasing attention in recent years, with automated writing evaluation (AWE) tools being at the forefront of this development (Grimes and Warschauer, 2010; Liu et al., 2022). AWE is an AI-powered technology that leverages Natural Language Processing (NLP) to evaluate and provide feedback on written texts. These tools are capable of identifying a wide range of linguistic features, such as grammar, vocabulary, coherence, and organization (Link et al., 2022). The immediate and personalized feedback provided by AWE can be useful for students who may not have regular access to writing tutors or instructors (Saricaoglu and Bilki, 2021). AWE can provide feedback on different types of writing tasks, including essays, research papers, and business reports, making it a versatile tool for a variety of educational contexts, such as language learning, academic writing, and workplace training. AWE tools are commonly integrated into Learning Management Systems (LMS) or writing platforms, and can be used as part of online writing courses or as standalone tools (Zhai and Ma, 2022).

The use of AWE in language learning has been found to improve writing skills, increase writing fluency, and enhance writing accuracy (Ranalli et al., 2017; Zhang and Hyland, 2018; Zhang, 2020; Ngo et al., 2022; Nunes et al., 2022). However, English as a foreign language (EFL) learners might face challenges in acquiring proficient writing skills due to limited opportunities for practicing writing and receiving feedback from experts (Hyland, 2007; Storch, 2011). In China, for instance, students often have limited opportunities to practice writing and receive individual feedback from teachers, who may have a large number of students and limited time. To address these challenges, AWE-based instructional programs have been occasionally employed in EFL classrooms in China (Tang and Rich, 2017). Despite this trend, there is little empirical research investigating the effectiveness of AWE on EFL writing skills in China. This study aims to address this gap by conducting a randomized controlled trial to examine the impact of AWE on the second language writing skills of Chinese EFL learners.

Anchored in the Social Cognitive Theory (SCT; Bandura, 2001), which emphasizes the roles of self-efficacy, observational learning, and self-regulation in learning and behavior change, this research investigates the effectiveness of AWE-based instructional programs on the development of writing skills among Chinese EFL learners. SCT underscores the intricate interplay between personal factors, environmental influences, and behavior within learning contexts. This theoretical framework posits that individuals learn through observing others and their interactions with the environment, enabling the development of self-efficacy beliefs that significantly influence motivation and performance. In the context of our study, the SCT offers a lens through which to comprehend how the AWE intervention, by delivering immediate and personalized feedback, could enhance learners' self-efficacy in writing tasks, thus potentially impacting their overall writing skills.

Against this backdrop, the purpose of this study is to investigate the effectiveness of AWE as an AI-powered system on second language writing skills of Chinese EFL learners. Specifically, this research aims to address the following research questions:

1. What is the impact of AWE on L2 writing skills of Chinese EFL learners?
2. Is the effect of AWE on L2 writing skills of Chinese EFL learners mediated by writing self-efficacy?
3. Does the effect of AWE on L2 writing skills of Chinese EFL learners differ across proficiency levels?

This study contributes to the existing literature by providing empirical evidence on the effectiveness of AWE-based instructional programs on the development of writing skills of Chinese EFL learners. The outcomes of this research can inform educators and researchers on the potential benefits of AWE-based instructional programs in EFL writing instruction, particularly in non-English speaking contexts.

2. Literature review

2.1. Theoretical framework

2.1.1. The technology acceptance model

Adoption theories aim to provide logic and explanation for people's intention to whether utilize an activity for the first time (Wallace and Sheetz, 2014). Concerning the technology use, the technology acceptance model (TAM) proposed by Davis (1989) has been a valid theoretical model to measure a one's degree of technological acceptance and evaluating the quality of e-learning. Evolved from the Theory of Reasoned Action (TRA), TAM tries to explain why users are willing or not to adopt technologies when performing a task (Wu and Chen, 2017). Technically speaking, TAM attempts to delve into the impact of technology on individuals' behavior (Moon and Kim, 2001). In fact, by focusing on two major factors, namely perceived usefulness and perceived ease of use, TAM explains user willingness to integrate a particular kind of technology (Abdullah et al., 2016). More specifically, Venkatesh and Davis (1996) proposed the final version of TAM composed of four underlying sub-constructs: perceived usefulness (i.e., how much a user believes that her/his job performance would be boosted while using a specific technology), perceived ease of use (i.e., how much a user believes that utilizing a specific technology would be unchallenging and effortless), external variables (i.e., factors which are at play when accepting a particular technology, such as user training, user engagement in design, technology characteristics, and the process of incorporating the technology), and behavioral intention (i.e., user's behavior towards utilizing a specific technology determined by her/his perceived usefulness and perceived ease of use) (Davis, 1989; Marangunic and Granic, 2015; Dizon, 2016). Due to its significant contribution, the Technology Acceptance Model (TAM) has frequently been acknowledged as the most influential and widely used theory for explaining an individual's adoption of information systems (Lee et al., 2003). Despite the wide investigation of the effectiveness and acceptance of computer based-technologies using TAM (e.g., Al-Azawei et al., 2017; Li et al., 2019; Fathi and Ebadi, 2020; Al-Azawei and Al-Azawi, 2021), fewer studies have employed TAM in the context of EFL learning and teaching. Also, the TAM model is still sorely underappreciated and insufficiently understood when applied to EFL field of study, and a comprehensive TAM model still needs to be investigated instantly.

2.2. Integrating AWE into L2 writing

In the 1960s the pioneer work regarding the automated scoring application was first developed trying to save teachers' time when scoring written texts and allowing teachers to provide feedback on learners' manuscripts (Parra and Calero, 2019). More importantly, given the improvements in artificial intelligence technology which has significantly contributed to the process of natural language and intelligent language system, the programs for automated grading have been upgraded and promoted since the 1990s (Liu et al., 2016). Consequently, numerous researchers have tried to develop computer-based applications and tools that can promote the writing skill and add value to scoring and feedback of it. As such, supported by the computer-mediated feedback technology, AWE is an ingenious technological tool that is implemented in various educational settings to provide evaluative feedback on learners' writing (Warschauer and Ware, 2006; Grimes and Warschauer, 2010). In fact, AWE is equipped with the kind of capacity that can constantly give qualitative and quantitative feedback on writing process by automatically scoring the text, analyzing the structure and creating a comprehensive evaluation of the text (Cotos, 2011; Li et al., 2014). The use of this technological tool is becoming increasingly common as a learning affordance in the learning process in various educational settings (Chen et al., 2009). Moreover, AWE is not only utilized for summative assessment in high-stakes writing tests but is also being effectively incorporated into classroom writing instruction.

According to Hassanzadeh and Fotoohnejad (2021), AWE plays a central role in the writing process, as it allows diagnostic and summative feedback to the learners. Furthermore, as Roscoe et al. (2017) asserted, AWE is a critical technological tool that saves teachers time when it comes to assessing writing, allows for more writing practice, and boost writing instruction. It is worth mentioning that one of the significant features of AWE tools is that they are interactive learning platforms. AWE tools often provide both build-in and customizable prompt for instructors to assign, as well as affording a diverse range of forms for the teacher to give comments on writing tasks (Palermo and Wilson, 2020). In addition, by using AWE, students are able to revise their manuscripts regarding the feedback they received from the source of the AWE tool, instructor, and peers (Geng and Razali, 2022).

Writing is often acknowledged as a demanding and intricate skill, particularly when it involves composing in a second language (Hashemian and Heidari, 2013; Marzban and Jalali, 2016; Hyland, 2019). This task becomes even more challenging for EFL learners, as acquiring writing proficiency in English poses difficulties not only for students but also for instructors (Cheung, 2016). As highlighted by Yu (2021), teaching writing skills, especially providing effective feedback on students' written work, can be a daunting endeavor for L2 teachers. However, writing is a skill that can be nurtured through consistent practice and timely feedback (Burstein et al., 2004; Fathi et al., 2020). When it comes to writing assessment, four distinct metrics that illuminate various facets of proficient written communication are usually employed (Polio, 1997; Uysal, 2010). The concept of task achievement, which gauges the extent to which a written piece fulfills given prompts or objectives, underscores the alignment between a writer's content and the prescribed context (Cumming, 2001). Coherence and cohesion, on the other hand, delve into the logical

organization and seamless connection of ideas within a text, ensuring its fluidity and accessibility to readers (Hyland, 2019). The lexical dimension, encompassing vocabulary selection and precision, significantly contributes to the depth and richness of expression (Nation and Nation, 2001). Finally, the aspect of grammatical accuracy, a pivotal component of effective communication, involves the meticulous application of language rules to convey meaning with clarity and precision (Bitchener and Ferris, 2012).

With respect to the field of language education (i.e., EFL), computer-based technologies have offered innovative trends of language instruction and language assessment which can be used for writing development and writing evaluation (Yousefifard and Fathi, 2021; Hsu and Lin, 2022; Parmaxi, 2023). As an appropriate technology to meet these needs, AWE automated serving can aid teachers with evaluating the texts and act as supporter which allows language learners to experience a sense of freedom and plan their own time to promote their motivation. Moreover, AWE is a technological tool which can generate timely and supportive feedback for EFL students in order to promote their writing process (Wang et al., 2013; Li et al., 2019; Ngo et al., 2022). As Jiang et al. (2020) demonstrated, AWE is an integral software that can significantly exert influence on L2 learners' writing skills. AWE computer-based programs can act as tools to evaluate EFL students' writing output and generate unique and individualized feedback (Jingxin and Razali, 2020; Fu et al., 2022). It is well-documented that the automated feedback provided by AWE in Second Language Acquisition (SLA) classrooms can offer significant benefits, such as writing longer texts, acquiring promoted machine scores, making fewer errors in essays, and boosting the rhetorical quality of writing texts (Li et al., 2015; Parra and Calero, 2019; Xu and Zhang, 2022). As put forward by Jingxin and Razali (2020), in L2 classrooms, AWE tools can offer authentic synchronous scores (i.e., holistic and analytic scores), as well as providing automated personalized diagnostic feedback on L2 students' manuscripts in various features of writing traits.

L2 teachers can integrate a variety of automatic feedback programs in classrooms to help them while teaching writing skills like wikis, MS Word computer software, and Grammar software among others (Zhang and Hyland, 2018; Stevenson and Phakiti, 2019). As one of the efficient automated feedback tools, Grammarly can be incorporated in L2 instruction to help learners and instructors in promoting EFL writing skills (Ebadi et al., 2023). Grammarly is an example of AWE that has gained particular attention as a practical tool in EFL classrooms. It can be integrated in L2 writing instruction to recognize structure deviations of texts, review spelling, punctuation, and check the originality to ensure that the text is mistake-free, clear, and polished (Ghufron and Rosyida, 2018; Barrot, 2022). This program, which is connected to the Internet, provides alternative words that are relevant if there are wrong words in the English language. Furthermore, Grammarly is incorporated into the Microsoft Word application which makes it a less demanding tool for learners to use to review deviations in English grammar with computer and suggests clarifications or samples of well-formed sentences and/or words. More importantly, the real-time writing feedback of Grammarly can assist EFL teachers to prevent writing deviations (Qassemzadeh and Soleimani, 2016). Grammarly contains an AI method which puts together deep learning and some approaches to natural language analysis in order to review grammatical constructs, phrases, paragraphs and written texts.

Previous research evidence indicates that AWE can greatly affect L2 students' writing skills (e.g., Liao, 2016a,b; Roscoe et al., 2017; Khoshnevisan, 2019; Jingxin and Razali, 2020; Lee, 2020; Tambunan et al., 2022; Waer, 2023). For instance, Liao (2016b) investigated the influence of the AWE-based approach in improving the writing accuracy of EFL students. To this end, 63 EFL learners took part in the study. Developing a 15-item questionnaire and a 12-question interview protocol, the findings indicated that AWE enhanced the writing accuracy of learners. In another study, Lee (2020) conducted a longitudinal study to explore the effects of AWE on Korean university learners' English writing competence. The perceptions towards their writing development which was acquired via interviews and journal entries were explored as well. Using a mixed-methods research design, the authors pointed to potential benefit of AWE, as it increased writing development and writing fluency of EFL students. In the context of Egypt, Waer (2023) explored the potential role of AWE in affecting EFL learners' writing process and grammatical competence. The findings revealed that AWE reduced writing apprehension and promoted the grammatical knowledge of participants. Also, Liao (2016b) examined the impact of AWE applications in shaping the writing improvement in an EFL context. The findings revealed that AWE facilitated the writing accuracy and writing development of EFL students. In their study, Saricaoğlu and Bilki (2021) investigated EFL students' written language under the influence of AWE. The findings indicated that EFL students' engagement with AWE significantly reduced their errors in writing and promoted their writing accuracy. Employing a mixed methods design, Wang et al. (2013) investigated the effect of integrating AWE in EFL university students' writing. Their outcomes revealed that AWE substantially enhanced EFL students' writing accuracy and promoted their autonomy awareness. With respect to other AWE tools, namely Grammarly, there are few studies which have examined the role of this tool in L2 writing (e.g., Khoshnevisan, 2019; Parra and Calero, 2019). Integrating Grammarly as an AWE tool, Khoshnevisan (2019) investigated the role of this software in developing and honing learners' writing skills. Gathering data from a sample of 12 students, the findings demonstrated that Grammarly contributed to participants' writing by motivating learners to develop their English writing competencies and produce more accurate essays. Moreover, it was found that Grammarly promoted English writing development by offering practical tips about grammar, vocabulary, and punctuation. Similarly, Parra and Calero (2019) found in their study that Grammarly was greatly conducive to EFL students' writing accuracy.

Despite the contributions of AWE tools (i.e., Grammarly) to EFL writing competencies, previous studies have mentioned some limitations to these technological programs. For instance, Stevenson and Phakiti (2014) demonstrated in their study that there is not much certainty regarding the positive effects of AWE on writing process, as it may not generate improvements in writing proficiency. The reason behind this may be attributed to the fact that computers-based technologies do not have the required judgement to evaluate those elements that are often associated with adequate writing, such as logic, clarity, accuracy, fluency, and relevancy. As Liao (2016a) demonstrated, AWE tools cannot perform imperfectly while addressing written language concerns (i.e., meaning, idea development, humor or irony, features of writing in which higher-order thinking is needed, quality of evidence, to name just a few). Therefore, due to its limitations and

drawbacks, AWE needs to be employed as a supplemental instrument rather than a substitute for instructor feedback.

Taken together, while many researches have mainly focused on the processes and perspectives of L2 teachers and learners, few researches have examined the role of AWE tools in affecting the L2 writing skills and competencies. Furthermore, most of the previous studies have investigated commercial AWE tools while neglecting others, namely Grammarly. More importantly, so far, the research regarding the integration of AWE in SLA domain is in nascent stages and little is known about this computer-based tool. In addition, to the best knowledge of the researcher, so far, few (if any) studies have surveyed the effects of AWE tools, namely Grammarly on EFL students' writing development. Hence, as an attempt to fill this research lacuna, the present study delved into the role of AWE in affecting L2 writing development and accuracy of EFL students, with a focus on the use of Grammarly.

2.3. The present study

The present study aimed to investigate the effectiveness of an Automated Writing Evaluation (AWE) tool on the second language writing skills of Chinese EFL learners. A randomized controlled trial (RCT) design was used, with participants being randomly assigned to either the experimental group or the control group. Both groups received 12 weeks of instruction, during which the treatment group underwent AWE-based instruction. In this instruction, participants utilized an AWE tool, Grammarly, to submit their written essays each week. The AWE tool provided immediate feedback on various aspects of writing, including grammar, spelling, vocabulary, and organization. Additionally, the treatment group attended weekly one-hour writing workshops designed to enhance their writing skills and incorporate the feedback from the AWE tool. In contrast, the control group received traditional writing instruction without the integration of AWE or the additional writing workshops. Based on the literature, it was hypothesized that the AWE-based writing instruction would lead to improvements in students' writing skills, as reflected by the four measures used in this study.

3. Methods

3.1. Participants

The AWE-based writing evaluation intervention was administered as an extracurricular program targeting intermediate EFL students in Mainland China. Informed written consent was obtained from all participants prior to their involvement in the intervention. The study cohort comprised 190 intermediate EFL students (60% female), all of whom were enrolled in one of four distinct writing courses hosted by different institutes offering the writing intervention. The participants' mean age was 21.5 years ($SD = 2.8$; range: 18–28 years).

To rigorously assess the efficacy of the writing intervention, we adopted a randomized controlled trial (RCT) design with repeated measures (Friedman et al., 2010). Initial measurements were conducted as pretests, seamlessly integrated into the first two sessions of the respective course. Subsequent posttest measurements were conducted during the final session of the course.

Within each institute, a control group was established, participating in a conventional writing course. Importantly, both the intervention and control courses were conducted simultaneously. The implementation of the AWE-based writing evaluation was overseen by a team of researchers collaborating with two proficient English teachers. It is essential to note that the AWE-based writing intervention remained consistent across all groups.

The randomization process was facilitated by bundling the two course options (AWE-based and conventional) under a single course-tandem, aptly named the *English Writing Course*. Enrollment into the course-tandem was exclusive, thus ensuring a controlled environment for the study. Post-enrollment, a blocked randomization technique was employed, utilizing computer-generated random numbers to allocate students to either the control or experimental groups. Through this approach, an equitable distribution of students was achieved across all participating institutes.

In total, 95 students were randomly assigned to the AWE-based writing intervention group (average age: $M = 21.6$, $SD = 2.9$; 60% female), while another 95 students were assigned to the traditional writing course group (average age: $M = 21.4$, $SD = 2.7$; 40% female). Following the study's completion, all students were invited to engage in the alternate course as a continuation of their learning process.

3.2. Instruments

3.2.1. Writing skills

In this study, two sample tasks from the International English Language Testing System (IELTS) were used to measure the writing skills of the participants. A pre-test task was administered to all participants before the intervention, serving as a baseline measure. Subsequently, a post-test task was given to both the experimental and control groups after the completion of the 12-week instructional period, which included the AWE-based instruction and traditional writing instruction without AWE, respectively. A pre-test task was administered before the intervention, while a post-test task was given after the AWE-based instruction. The writing performance of the participants was assessed using an analytic essay scoring scale based on the IELTS rubric.

The IELTS rubric, renowned for its reliability, is extensively employed for assessing writing abilities within second language contexts. This rubric employs a range of scores, typically from 1 to 9, to evaluate distinct descriptors across various dimensions of writing, such as task accomplishment, coherence and cohesion, lexical richness, and grammatical precision. Each criterion encompasses specific descriptors that correspond to different levels of proficiency, and these descriptors are scored individually within the established score range.

The final score derived from the IELTS rubric is calculated as the mean score of the descriptors. In this method, each descriptor's score is assigned a weight based on its significance within the overall writing competence. The individual scores for task achievement, coherence and cohesion, lexical resource, and grammatical accuracy are averaged to determine the participant's final writing proficiency score. This approach provides a comprehensive and nuanced assessment of the participants' writing skills, accounting for their performance across a spectrum of criteria.

The selection of the IELTS rubric for the analytic essay scoring scale was based on its comprehensive nature and established reliability and validity in assessing writing skills. By employing the IELTS rubric, this study ensured a standardized and consistent evaluation of participants' writing performance, enabling a reliable comparison of their progress and the impact of the AWE intervention. To ensure the consistency of the scoring process, two independent raters were recruited, and inter-rater reliability was calculated using Cohen's Kappa, which was reported to be 0.82.

3.2.2. Global English proficiency

To evaluate the participants' general English language proficiency and ensure their comparability, the Oxford Placement Test (OPT) developed by Allan (2004) was employed. The OPT is a versatile assessment tool that accurately determines the appropriate proficiency level for English learners, evaluating dimensions such as vocabulary, grammar, listening comprehension, and reading skills. The internal consistency of the OPT, assessed using Cronbach's alpha, yielded a reliability coefficient of 0.83 in this study, indicating a high level of internal reliability.

To enhance the comparability of the OPT scores with the IELTS rubric, the total scores obtained from the OPT were transformed onto a 0–9 scale. This conversion was undertaken to align the OPT scores with the scoring scale familiarly associated with the IELTS rubric. This approach allowed for a consistent interpretation of participants' language proficiency across both assessments, providing a unified framework for evaluating their language skills.

3.2.3. Writing self-efficacy

To measure the writing self-efficacy of L2 students, the scale developed by Han and Hiver (2018) was utilized. This scale consisted of seven items designed to assess students' beliefs and assurance in their writing abilities. The questionnaire adopted a 5-point Likert scale format, ranging from 1 (strongly disagree) to 5 (strongly agree). The internal consistency of the scale, assessed using Cronbach's Alpha, was found to be 0.78 in the present investigation.

3.2.4. Procedure

The experimental intervention in this study aimed to enhance the L2 writing competencies of Chinese EFL learners through the use of an AWE tool. The AWE tool was provided by Grammarly and was used by the students to submit a written essay in English every week for a period of 12 weeks. The tool provided immediate feedback on various aspects of writing, including grammar, spelling, vocabulary, and organization. The feedback was given in the form of suggested corrections and explanations, which the students were encouraged to review and incorporate into their subsequent writing.

In addition to the AWE tool, the students in the experimental group received a weekly one-hour writing workshop that focused on developing their writing skills and providing additional opportunities for practice. The writing workshop was designed to complement the AWE tool by giving learners the individualized feedback on their writing, as well as guidance on how to improve their writing skills. The workshop covered various aspects of writing, including grammar, vocabulary, sentence structure, and organization.

On the other hand, the control group in this study received traditional writing instruction without the use of an AWE tool. The students in the control group were asked to write an essay in English

every week for a period of 8 weeks, which were graded by the instructor based on a rubric that evaluated various aspects of writing, including grammar, spelling, vocabulary, and organization. The students in the control group also received a weekly one-hour writing workshop that was similar in content and structure to the workshops provided to the experimental group. However, the writing workshops in the control group did not include the use of an AWE tool.

Overall, the experimental intervention in this study aimed to improve the second language writing skills of Chinese EFL learners by providing them with immediate feedback on their writing using an AWE tool and additional opportunities for practice through weekly writing workshops. The control group, on the other hand, aimed to improve the second language writing skills of Chinese EFL learners through traditional writing instruction without the use of an AWE tool. The effectiveness of these two approaches was compared to determine the impact of AWE on the second language writing skills of Chinese EFL learners.

3.3. Ensuring treatment fidelity

In order to ensure the validity of the results, treatment fidelity was closely monitored across all groups. To achieve this, a guideline was developed to provide the instruction to two pilot groups of 10 and 12 EFL students prior to the actual study. The teaching materials and course content were standardized for all groups and given in the same order. In addition, the pretest and all trainings were conducted by the research team to ensure consistency and fidelity to the experimental design. By implementing these measures, the study ensured that the intervention was delivered as intended and that any differences observed between the experimental and control groups could be confidently attributed to the use of the AWE tool. This approach is consistent with prior research on treatment fidelity (Graham and Harris, 2014) and strengthens the internal validity of the study.

3.4. Analysis

To evaluate the effectiveness of the AWE-based instruction, four measures including task achievement, coherence and cohesion, lexicon, and grammatical accuracy were used. To enhance the accuracy of the regression coefficients and mitigate potential biases resulting from between-group differences at the study's outset (Cohen et al., 2003), control variables were incorporated. The first control variable was global English proficiency, which was measured using the OPT. The second control variable was writing self-efficacy, which was included due to its potential impact on writing performance. To further explore the impact of the AWE-based instruction, the interaction term of the course and pretest score was included as an additional predictor variable. This allowed us to assess the differential effects of the intervention for EFL students with low versus high pretest scores on the dependent variable.

To ensure that the training conditions did not differ significantly at the outset of the research, two-tailed *t*-tests were performed to examine the pretest measures for all dependent and control variables. The baseline equivalence was examined for key characteristics. The dependent variables included posttest measurements for each of the four sub-scales (i.e., Task Achievement, Coherence and cohesion,

Lexicon, Grammatical accuracy) of writing skills. To evaluate the intervention's efficacy, multiple linear regression analyses were employed using Mplus Version 7 (Muthén and Muthén, 1998–2012) with maximum likelihood robust estimation (MLR). The predictors were entered simultaneously into the multiple linear regression model. This approach allowed us to examine the collective impact of all predictors on the dependent variable, writing skills. By including all predictors together, we aimed to understand how their combined effects contribute to explaining the variance in writing skills among the participants. The percentage of missing data ranged from 2 to 6%, and there was no significant differential drop-out between the treatment and control groups [$\chi^2(1, 190)=1.08; p=0.299$]. Significance tests were one-tailed, with a significance level (α) set at 0.05. Hypotheses were formulated in a directed manner to examine the training effects.

The full-information maximum likelihood (FIML) estimator was used to handle missing data, assuming that the missing data were missing at random (Enders, 2010). FIML analysis is a statistical approach that utilizes all available data to estimate parameters and standard errors (Buhi et al., 2008). Prior to the analyses, continuous variables were standardized. The experimental and control groups were represented as binary variables, with a value of 1 assigned to the experimental group and 0 to the control group. The magnitude of the intervention impact was assessed by comparing the standardized mean differences (Hedges, 2007). As no similar studies were found, the widely accepted classification of effect sizes was employed: small ($d=0.20$), medium ($d=0.50$), and large ($d=0.80$) (Cohen, 1992). Since treatment effects were assessed across four dependent variables, the Benjamini-Hochberg procedure (Benjamini and Hochberg, 1995) was employed to control for multiple testing, and adjusted *p*-values were reported.

4. Results

Table 1 displays the means and standard deviations for the experimental and control groups in the pre- and post-tests. The experimental group received Automated Writing Evaluation (AWE) intervention while the control group received traditional writing instruction. The table shows the scores for task achievement, coherence and cohesion, lexicon, grammatical accuracy, global English proficiency (measured by OPT), and writing self-efficacy. The pretest means for both groups were similar for all measures, and no significant differences were found. However, at the posttest, the experimental group showed higher mean scores in all measures than the control group. The missing data ranged from 2 to 6% across both groups, with the higher missing rate resulting from students' absence at posttest. The missing data in this study were determined to be missing at random. The interclass correlation coefficients (ICC) for all measures were above 0.70, indicating acceptable levels of reliability.

Table 2 presents the correlations at the pretest (below diagonal) and posttest (above diagonal) in the study. The table shows the Pearson's correlation coefficients between the four variables measured in the study.

The table indicates that there is a statistically significant positive correlation between task achievement and coherence and cohesion at both the pretest ($r=0.24, p<0.05$) and the posttest ($r=0.43, p<0.01$).

TABLE 1 Means and standard deviations for each group in pre- and post-tests.

	Pre-test				MIS	Post-test				MIS	ICC
	Experimental G		Control G			Experimental G		Control G			
	<i>M</i>	<i>SD</i>	<i>M</i>	<i>SD</i>		<i>M</i>	<i>SD</i>	<i>M</i>	<i>SD</i>		
Task	4.26	0.83	4.12	0.61	6	5.42	0.69	4.78	0.83	11	0.82
Coherence	4.12	0.69	4.20	0.72	6	5.28	0.74	5.12	0.59	11	0.79
Lexicon	5.14	0.92	4.98	0.49	6	6.02	0.85	5.36	0.97	11	0.71
Accuracy	3.86	0.68	3.97	0.82	6	5.76	0.92	4.68	0.68	11	0.90
Global Eng	5.46	1.24	5.33	0.89	3						
WSE	3.57	0.72	3.71	0.92	3						

Task, task achievement; Coherence, coherence and cohesion; Accuracy, grammatical accuracy; Global Eng, Global English proficiency (OPT); WSE, writing self-efficacy.

TABLE 2 Correlations among the constructs at the pretest (below diagonal) and the posttest (above diagonal).

	1	2	3	4
(1) Task achievement	–	0.22*	0.37**	0.31**
(2) Coherence and Cohesion	0.24*	–	0.43**	0.38**
(3) Lexicon	0.27*	0.22*	–	0.41**
(4) Grammatical accuracy	0.32**	0.26*	0.33**	–

* $p < 0.05$; ** $p < 0.01$.

There is also a significant positive correlation between task achievement and lexicon at the pretest ($r = 0.27, p < 0.05$) and the posttest ($r = 0.41, p < 0.01$). Similarly, there is a significant positive correlation between coherence and cohesion and lexicon at both the pretest ($r = 0.22, p < 0.05$) and the posttest ($r = 0.38, p < 0.01$).

Furthermore, there is a significant positive correlation between grammatical accuracy and task achievement at the pretest ($r = 0.32, p < 0.01$) and posttest ($r = 0.33, p < 0.01$). There is also a significant positive correlation between grammatical accuracy and coherence and cohesion at the pretest ($r = 0.26, p < 0.05$) but not at the posttest.

Table 3 reports the results of an analysis of the effects of AWE-based instruction on writing skills (posttest) in terms of task achievement, coherence and cohesion, lexicon, and grammatical accuracy.

In terms of the intervention effect, the results show that AWE-based instruction has a significant positive effect on task achievement ($B = 0.38, SE = 0.27, p = 0.044$), coherence and cohesion ($B = 0.46, SE = 0.32, p = 0.036$), lexicon ($B = 0.55, SE = 0.31, p = 0.009$), and grammatical accuracy ($B = 0.74, SE = 0.29, p = 0.003$).

The results also show that the pretest score is a significant predictor of task achievement ($B = 0.26, SE = 0.19, p = 0.197$), coherence and cohesion ($B = 0.34, SE = 0.26, p = 0.046$), and lexicon ($B = 0.33, SE = 0.24, p = 0.245$), but not of grammatical accuracy ($B = 0.19, SE = 0.13, p = 0.573$).

The interaction effect between the intervention and pretest score is not significant for task achievement ($B = 0.29, SE = 0.24, p = 0.621$) and coherence and cohesion ($B = 0.21, SE = 0.17, p = 0.263$), but is significant for lexicon ($B = -0.39, SE = 0.24, p = 0.048$) indicating that the effect of the intervention on Lexicon is weaker for participants who had higher pretest scores.

Global English proficiency is not a significant predictor of any of the outcome variables. However, writing self-efficacy (WSE) is a

significant predictor of task achievement ($B = 0.39, SE = 0.19, p = 0.024$) and Coherence and cohesion ($B = 0.52, SE = 0.30, p = 0.013$), but not of Lexicon ($B = 0.26, SE = 0.19, p = 0.427$) or grammatical accuracy ($B = 0.24, SE = 0.17, p = 0.092$).

The explained variance (R^2) shows that the AWE-based instruction accounts for 31% of the variance in task achievement, 28% in coherence and cohesion, 35% in lexicon, and 46% in grammatical accuracy.

Finally, the omnibus test for the overall model was statistically significant ($F = 17.12, p < 0.001$), indicating that the combination of predictors significantly improved the fit of the model compared to a null model. This suggests that the included predictors collectively contribute to the prediction of writing skills among the participants.

5. Discussion

The present study aimed to investigate the effectiveness of AWE-based instruction on the second language writing skills of Chinese EFL learners. More specifically, the researchers examined the effects of AWE-based instruction on task achievement, coherence and cohesion, lexicon, and grammatical accuracy, while also considering global English proficiency, writing self-efficacy, and pre-test scores as control variables. The results indicate that AWE-based instruction had a significant positive effect on L2 writing skills.

The positive effect of AWE-based instruction on task achievement is consistent with previous studies that have highlighted the role of automated feedback in enhancing learners' ability to meet specific writing task requirements effectively (Liao, 2016a; Jiang et al., 2020; Barrot, 2022; Jiang and Yu, 2022). Via providing immediate and targeted feedback, AWE systems enable learners to identify and address gaps in task achievement, leading to improved performance. Similarly, the positive impact of AWE-based instruction on coherence and cohesion supports previous research highlighting the role of technology in promoting cohesive and well-structured writing (Tuzi, 2004; Li et al., 2019; Kessler, 2020; Rahimi and Fathi, 2022). AWE systems can assist learners in identifying and rectifying issues related to paragraph organization, sentence connections, and the overall flow of ideas (Cotos, 2011). The improvement observed in lexicon and grammatical accuracy can be attributed to the automated features of AWE systems, which enable learners to receive detailed feedback on vocabulary use and grammatical errors (Zhang and Hyland, 2018; Fu et al., 2022). The immediate feedback provided by AWE systems allows learners to

TABLE 3 AWE-based instruction effects on the writing skills (posttest).

	Task achievement			Coherence and cohesion			Lexicon			Grammatical accuracy		
	<i>B</i>	<i>SE</i>	<i>p</i>	<i>B</i>	<i>SE</i>	<i>p</i>	<i>B</i>	<i>SE</i>	<i>p</i>	<i>B</i>	<i>SE</i>	<i>p</i>
Intervention	0.38	0.27	0.044	0.46	0.32	0.036	0.55	0.31	0.009	0.74	0.29	0.003
Pretest score	0.26	0.19	0.197	0.34	0.26	0.046	0.33	0.24	0.245	0.19	0.13	0.573
Intervention × Pretest score	0.29	0.24	0.621	0.21	0.17	0.263	−0.39	0.24	0.048	−0.30	0.26	0.140
Global English	0.75	0.32	1	−0.32	0.24	0.186	−0.19	0.15	0.792	0.16	0.13	0.954
WSE	0.39	0.19	0.024	0.52	0.30	0.013	0.26	0.19	0.427	0.24	0.17	0.092
Explained variance (<i>R</i> ²)	0.31			0.28			0.35			0.46		

identify and correct lexical and grammatical issues, leading to enhanced language accuracy (Ranalli, 2018; Zhang, 2020).

Taken together, the findings of this study provide empirical evidence supporting the positive impact of AWE-based instruction on multiple components of writing skills among Chinese EFL learners. These results align with previous studies that have also reported the effectiveness of AWE tools on enhancing second language writing skills (e.g., Liao, 2016a,b; Ranalli et al., 2017; Li et al., 2019; Hassanzadeh and Fotoohnejad, 2021; Barrot, 2022; Fu et al., 2022; Ebadi et al., 2023).

One likely reason for the effectiveness of AWE-based instruction is that it provides immediate and personalized feedback to learners, allowing them to identify and correct their errors in real-time. This feature is especially beneficial for low-proficiency learners who may struggle with self-correction and need more guidance in their writing process.

This finding aligns with previous research on the benefits of technology in language learning, particularly in improving writing skills (Tuzi, 2004; Stapleton and Radia, 2010; Kessler, 2020; Fathi and Rahimi, 2022). This study is anchored in the Social Cognitive Theory (SCT) which suggests that people learn by observing and understanding the consequences of their actions (Bandura, 2003). In the context of language learning, AWE offers immediate feedback to students on their writing, allowing them to recognize the outcomes of their writing strategies and make the necessary adjustments (Bandura, 2001). Therefore, the findings of this study provide empirical evidence that supports SCT's belief that feedback is a fundamental aspect of the learning process. Nevertheless, it is important to note that AWE should not replace human feedback and evaluation entirely. Warschauer and Healey (1998) suggested that technology should supplement and support human instruction rather than replacing it completely. Thus, AWE should be employed in tandem with teacher feedback and instruction to provide a well-rounded approach to writing instruction in EFL environments.

Furthermore, the results indicate that pre-test scores significantly predicted task achievement, coherence and cohesion, and lexicon. Although this finding suggests that learners' initial proficiency levels exert influence on their subsequent writing performance, it is crucial to clarify the steps taken to mitigate the potential impact of participants' initial differences on the observed writing skill differences. To address this concern, the study employed a randomized controlled trial design, which ensured that

participants were assigned to the experimental and control groups randomly. This random assignment aimed to distribute any potential initial skill disparities evenly across both groups. Additionally, the study employed repeated measures, allowing for within-subject comparisons over time, effectively controlling for individual differences. This design choice aimed to provide a comprehensive understanding of the intervention's impact by examining how each participant's skills evolved relative to their own baseline. The absence of a significant interaction effect between the intervention and pre-test scores for task achievement and coherence and cohesion indicates that the impact of AWE-based instruction remained consistent across varying proficiency levels in these aspects. However, the significant interaction effect for lexicon indicates that the intervention's effect on lexical improvement was weaker for participants with higher pre-test scores. This finding suggests that learners with higher initial lexical proficiency may have had less room for improvement in this specific aspect. This is consistent with previous research on the effectiveness of technology-enhanced language learning for low-proficiency learners (Huang et al., 2017; Zhang et al., 2022). These studies suggest that technology-enhanced language learning can provide more individualized and personalized instruction (Golonka et al., 2014) that caters to the specific needs of low-proficiency learners, thus leading to more effective learning outcomes.

On the other hand, writing self-efficacy was found to be a significant predictor of task achievement and coherence and cohesion. This finding highlights the importance of learners' beliefs and confidence in their writing abilities. Higher levels of writing self-efficacy may contribute to increased motivation and effort invested in writing tasks, leading to improved performance in specific writing skill components. These findings are consistent with theoretical frameworks emphasizing the role of self-efficacy beliefs in influencing learners' engagement and success in writing activities (Lee and Evans, 2019; Golkarparvar and Khafi, 2021; Tsao, 2021).

Moreover, this finding is consistent with the Sociocultural Theory, which suggests that learning is a social and cultural process (Vygotsky, 1978). This theory maintains that learners' language learning experiences are influenced by their social and cultural background, including their beliefs, prior knowledge, and experiences. Individuals with lower proficiency levels may have limited exposure to the target language and culture, which can restrict their language learning

opportunities. Technology-aided language learning can provide more tailored and personalized instruction, which can assist low-proficiency learners in overcoming these obstacles and promoting their language learning.

Taken together, this study indicates that AWE can be a useful tool in enhancing the writing skills of Chinese EFL learners, particularly those with lower proficiency levels. The results of this study align with the existing literature on technology-enhanced language learning and the significance of feedback in the learning process. The study's implications for EFL educators and researchers are to consider integrating AWE into their teaching and learning practices. Nonetheless, this study's limitations include using a single measure for writing skills and the need for further research on the long-term effectiveness of AWE on language learning outcomes. This study contributes to the current body of research on technology-enhanced language learning's potential to improve language skills and provides valuable insights into the use of AWE in EFL settings.

6. Conclusion and implications

In this study, the researchers probed the utility of an AWE-based instructional program on the writing skills of Chinese EFL students. The outcomes showed that the program was successful in improving L2 writing skills, with greater benefits for low-proficiency students. These findings have significant implications for second language writing instruction, suggesting that educators should incorporate AWE-based tools like Grammarly into their teaching practice. By providing instant feedback, AWE can support self-directed learning and personalized instruction, helping learners develop a comprehensive set of writing skills that includes more than just grammatical accuracy. Therefore, AWE can complement traditional writing instruction and improve learners' overall writing abilities.

In addition, this study has important implications for curriculum design and assessment. The inclusion of AWE in the curriculum can help students become more familiar with AI-based tools, which can prepare them for academic and professional contexts where such tools are commonly used. Additionally, AWE can serve as an objective assessment tool that can reliably measure student progress in writing skills, providing teachers with valuable feedback. As technology use in language education continues to expand, it is crucial to explore the potential advantages and disadvantages of various tools and approaches. The use of AWE-based instructional programs can offer a more effective and objective way of assessing and improving L2 writing skills, especially in situations where face-to-face instruction is not possible. Overall, this study demonstrates the potential of AWE-based instructional programs to improve second language writing skills and suggests that integrating AI-based tools can be a promising approach to enhance the effectiveness of second language writing instruction.

While the results of this study suggest that AWE-based instructional programs can be effective in improving second language writing skills, there are several limitations that should be considered. One limitation is that the study was conducted with

a sample of Chinese EFL learners only, which limits the generalizability of the findings to other EFL contexts. To address this limitation, future research should replicate the study with different populations. Another limitation is that the study only examined short-term effects of AWE-based instruction, and it is unclear whether these effects would persist over time. Therefore, future studies should investigate the long-term effects of AWE-based instructional programs on second language writing skills. Furthermore, the study did not explore the attitudes of learners towards using AWE technology in writing instruction. Understanding students' acceptance and engagement with the technology is important to determine the effectiveness of AWE-based instruction. Therefore, future research should investigate learners' attitudes and perceptions towards AWE-based instruction. Lastly, the study did not examine the effects of AWE-based instruction on other aspects of writing, such as discourse organization and rhetorical strategies. Future studies should explore whether AWE-based instruction can improve these aspects of second language writing as well.

Data availability statement

The raw data supporting the conclusions of this article will be made available by the authors, without undue reservation. Requests to access these datasets should be directed to PW, peiqi469@163.com.

Ethics statement

The studies involving humans were approved by the School of Foreign Languages, Tangshan Normal University, Tangshan, Hebei, China. The studies were conducted in accordance with the local legislation and institutional requirements. The participants provided their written informed consent to participate in this study.

Author contributions

All the authors equally contributed to completing this project.

Conflict of interest

The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Publisher's note

All claims expressed in this article are solely those of the authors and do not necessarily represent those of their affiliated organizations, or those of the publisher, the editors and the reviewers. Any product that may be evaluated in this article, or claim that may be made by its manufacturer, is not guaranteed or endorsed by the publisher.

References

- Abdullah, F., Ward, R., and Ahmed, E. (2016). Investigating the influence of the most commonly used external variables of TAM on students' perceived ease of use (PEOU) and perceived usefulness (PU) of e-portfolios. *Comput. Hum. Behav.* 63, 75–90. doi: 10.1016/j.chb.2016.05.014
- Al-Azawei, A., Parslow, P., and Lundqvist, K. (2017). Investigating the effect of learning styles in a blended e-learning system: an extension of the technology acceptance model (TAM). *Australas. J. Educ. Technol.* 33, 1–23. doi: 10.14742/ajet.2741
- Al-Azawei, A., and Al-Azawi, R. (2021). Evaluating Facebook success in Iraq: An extension of the DeLone and McLean's model of information systems success (ISS). In *J. Phys. Conf. Ser.* IOP Publishing, 1804:012114. doi: 10.1088/1742-6596/1804/1/012114
- Allan, D. (2004). *Oxford placement test*. Oxford: Oxford University Press.
- Bandura, A. (2001). Social cognitive theory: an agentic perspective. *Annu. Rev. Psychol.* 52, 1–26. doi: 10.1146/annurev.psych.52.1.1
- Bandura, A. (2003). "Social cognitive theory for personal and social change by enabling media" in *Entertainment—education and social change: History, research, and practice*. eds. A. Singhal, M. J. Cody, E. M. Rogers and M. Sabido (Mahwah, NJ: Lawrence Erlbaum), 75–96.
- Barrot, J. S. (2022). Integrating technology into ESL/EFL writing through Grammarly. *RELJ* 53, 764–768. doi: 10.1177/0033688220966632
- Benjamini, Y., and Hochberg, Y. (1995). Controlling the false discovery rate: a practical and powerful approach to multiple testing. *J. R. Stat. Soc. Ser. B* 57, 289–300. doi: 10.1111/j.2517-6161.1995.tb02031.x
- Bitchenor, J., and Ferris, D. R. (2012). *Written corrective feedback in second language acquisition and writing*. New York, NY: Routledge.
- Buhi, E. R., Goodson, P., and Neilands, T. B. (2008). Out of sight, not out of mind: strategies for handling missing data. *Am. J. Health Behav.* 32, 83–92. doi: 10.5993/AJHB.32.1.8
- Burstein, J., Chodorow, M., and Leacock, C. (2004). Automated essay evaluation: the criterion online writing service. *AI Mag.* 25:27. doi: 10.1609/aimag.v25i3.1774
- Chen, H. J., Chiu, T. L., and Liao, P. (2009). Analyzing the grammar feedback of two automated writing evaluation systems: my access and criterion. *English Teaching & Learning* 33, 1–43.
- Cheung, Y. L. (2016). Teaching writing. In *English language teaching today*. eds. W. A. Renandya and H. P. Widodo. Springer, 179–194.
- Cohen, J. (1992). A power primer. *Psychol. Bull.* 112, 155–159. doi: 10.1037/0033-2909.112.1.155
- Cohen, J., Cohen, P., West, S. G., and Aiken, L. S. (2003). *Applied multiple regression/correlation analysis for the behavioral sciences (3rd Edn.)*. Mahwah, NJ: L. Erlbaum Associates.
- Cotos, E. (2011). Potential of automated writing evaluation feedback. *CALICO J.* 28, 420–459. doi: 10.11139/cj.28.2.420-459
- Cumming, A. (2001). ESL/EFL instructors' practices for writing assessment: specific purposes or general purposes? *Lang. Test.* 18, 207–224. doi: 10.1191/026553201678777086
- Davis, F. D. (1989). Perceived usefulness, perceived ease of use, and user acceptance of information technology. *MIS Q.* 13:319. doi: 10.2307/249008
- Dizon, G. (2016). Measuring Japanese EFL student perceptions of internet-based tests with the technology acceptance model. *TESL-Ej* 20:n2
- Ebadi, S., Gholami, M., and Vakili, S. (2023). Investigating the effects of using Grammarly in EFL writing: the case of articles. *Comput. Sch.* 40, 85–105. doi: 10.1080/07380569.2022.2150067
- Enders, C. K. (2010). *Applied missing data analysis. Methodology in the social sciences*. New York, NY: Guilford Press.
- Fathi, J., Derakhshan, A., and Safdari, M. (2020). The impact of portfolio-based writing instruction on writing performance and anxiety of EFL students. *Pol. Psychol. Bull.* 51, 226–235. doi: 10.24425/ppb.2020.134729
- Fathi, J., and Ebadi, S. (2020). Exploring EFL pre-service teachers' adoption of technology in a CALL program: obstacles, motivators, and maintenance. *Educ. Inf. Technol.* 25, 3897–3917. doi: 10.1007/s10639-020-10146-y
- Fathi, J., and Rahimi, M. (2022). Electronic writing portfolio in a collaborative writing environment: its impact on EFL students' writing performance. *Comput. Assist. Lang. Learn.* 1-39, 1–39. doi: 10.1080/09588221.2022.2097697
- Friedman, L. M., Furberg, C., and DeMets, D. L. (2010). *Fundamentals of clinical trials (4th ed.)*. New York, NY: Springer. doi: 10.1007/978-3-319-18539-2
- Fu, Q. K., Zou, D., Xie, H., and Cheng, G. (2022). A review of AWE feedback: types, learning outcomes, and implications. *Comput. Assist. Lang. Learn.* 1-43, 1–43. doi: 10.1080/09588221.2022.2033787
- Geng, J., and Razali, A. B. (2022). Effectiveness of the automated writing evaluation program on improving undergraduates' writing performance. *Engl. Lang. Teach.* 15, 49–60. doi: 10.5539/elt.v15n7p49
- Ghufuron, M. A., and Rosyida, F. (2018). The role of Grammarly in assessing English as a foreign language (EFL) writing. *Ling. Cult.* 12, 395–403. doi: 10.21512/lc.v12i4.4582
- Golonka, E. M., Bowles, A. R., Frank, V. M., Richardson, D. L., and Freynik, S. (2014). Technologies for foreign language learning: a review of technology types and their effectiveness. *Comput. Assist. Lang. Learn.* 27, 70–105. doi: 10.1080/09588221.2012.700315
- Golparvar, S. E., and Khafi, A. (2021). The role of L2 writing self-efficacy in integrated writing strategy use and performance. *Assess. Writ.* 47:100504. doi: 10.1016/j.asw.2020.100504
- Graham, S., and Harris, K. R. (2014). Conducting high quality writing intervention research: twelve recommendations. *J. Writ. Res.* 6, 89–123. doi: 10.17239/jowr-2014.06.02.1
- Grimes, D., and Warschauer, M. (2010). Utility in a fallible tool: a multi-site case study of automated writing evaluation. *J. Technol. Learn. Assess.* 8, 4–42
- Han, J., and Hiver, P. (2018). Genre-based L2 writing instruction and writing-specific psychological factors: the dynamics of change. *J. Second. Lang. Writ.* 40, 44–59. doi: 10.1016/j.jslw.2018.03.001
- Hashemian, M., and Heidari, A. (2013). The relationship between L2 learners' motivation/attitude and success in L2 writing. *Procedia Soc. Behav. Sci.* 70, 476–489. doi: 10.1016/j.sbspro.2013.01.085
- Hassanzadeh, M., and Fotoohnejad, S. (2021). Implementing an automated feedback program for a foreign language writing course: a learner-centric study: implementing an AWE tool in a L2 class. *J. Comput. Assist. Learn.* 37, 1494–1507. doi: 10.1111/jcal.12587
- Hedges, L. V. (2007). Effect sizes in cluster-randomized designs. *J. Educ. Behav. Stat.* 32, 341–370. doi: 10.3102/1076998606298043
- Hsu, H. T., and Lin, C. C. (2022). Extending the technology acceptance model of college learners' mobile-assisted language learning by incorporating psychological constructs. *Br. J. Educ. Technol.* 53, 286–306. doi: 10.1111/bjet.13165
- Huang, Y. Y., Liu, C. C., Wang, Y., Tsai, C. C., and Lin, H. M. (2017). Student engagement in long-term collaborative EFL storytelling activities: an analysis of learners with English proficiency differences. *J. Educ. Technol. Soc.* 20, 95–109.
- Hyland, K. (2007). Genre pedagogy: language, literacy and L2 writing instruction. *J. Second. Lang. Writ.* 16, 148–164. doi: 10.1016/j.jslw.2007.07.005
- Hyland, K. (2019). *Second language writing*. Cambridge: Cambridge university press.
- Jiang, L., and Yu, S. (2022). Appropriating automated feedback in L2 writing: experiences of Chinese EFL student writers. *Comput. Assist. Lang. Learn.* 35, 1329–1353. doi: 10.1080/09588221.2020.1799824
- Jiang, L., Yu, S., and Wang, C. (2020). Second language writing instructors' feedback practice in response to automated writing evaluation: a sociocultural perspective. *System* 93:102302. doi: 10.1016/j.system.2020.102302
- Jingxin, G., and Razali, A. B. (2020). Tapping the potential of pigai automated writing evaluation (AWE) program to give feedback on EFL writing. *Univ. J. Educ. Res.* 8, 8334–8343. doi: 10.13189/ujer.2020.082638
- Kessler, M. (2020). Technology-mediated writing: exploring incoming graduate students' L2 writing strategies with activity theory. *Comput. Compos.* 55:102542. doi: 10.1016/j.compcom.2020.102542
- Khoshnevisan, B. (2019). The affordances and constraints of automatic writing evaluation (AWE) tools: a case for grammarly. *Artesol EFL J.* 2, 12–25.
- Lee, Y. J. (2020). The long-term effect of automated writing evaluation feedback on writing development. *Eng. Teach.* 75, 67–92. doi: 10.15858/engtea.75.1.202003.67
- Lee, M. K., and Evans, M. (2019). Investigating the operating mechanisms of the sources of L2 writing self-efficacy at the stages of giving and receiving peer feedback. *Mod. Lang. J.* 103, 831–847. doi: 10.1111/modl.12598
- Lee, Y., Kozar, K. A., and Larsen, K. R. T. (2003). The technology acceptance model: past, present, and future. *Commun. Assoc. Inf. Syst.* 12, 752–780. doi: 10.17705/1CAIS.01250
- Li, J., Link, S., and Hegelheimer, V. (2015). Rethinking the role of automated writing evaluation (AWE) feedback in ESL writing instruction. *J. Second. Lang. Writ.* 27, 1–18. doi: 10.1016/j.jslw.2014.10.004
- Li, Z., Link, S., Ma, H., Yang, H., and Hegelheimer, V. (2014). The role of automated writing evaluation holistic scores in the ESL classroom. *System* 44, 66–78. doi: 10.1016/j.system.2014.02.007
- Li, R., Meng, Z., Tian, M., Zhang, Z., Ni, C., and Xiao, W. (2019). Examining EFL learners' individual antecedents on the adoption of automated writing evaluation in China. *Comput. Assist. Lang. Learn.* 32, 784–804. doi: 10.1080/09588221.2018.1540433
- Liao, H. C. (2016a). Using automated writing evaluation to reduce grammar errors in writing. *ELT J.* 70, 308–319. doi: 10.1093/elt/ccv058
- Liao, H. C. (2016b). Enhancing the grammatical accuracy of EFL writing by using an AWE-assisted process approach. *System* 62, 77–92. doi: 10.1016/j.system.2016.02.007

- Link, S., Mehrzad, M., and Rahimi, M. (2022). Impact of automated writing evaluation on teacher feedback, student revision, and writing improvement. *Comput. Assist. Lang. Learn.* 35, 605–634. doi: 10.1080/09588221.2020.1743323
- Liu, G. Z., Rahimi, M., and Fathi, J. (2022). Flipping writing metacognitive strategies and writing skills in an English as a foreign language collaborative writing context: a mixed-methods study. *J. Comput. Assist. Learn.* 38, 1730–1751. doi: 10.1111/jcal.12707
- Liu, O. L., Rios, J. A., Heilman, M., Gerard, L., and Linn, M. C. (2016). Validation of automated scoring of science assessments. *J. Res. Sci. Teach.* 53, 215–233. doi: 10.1002/tea.21299
- Marangunic, N., and Granic, A. (2015). Technology acceptance model: a literature review from 1986 to 2013. *Univ. Access Inf. Soc.* 14, 81–95. doi: 10.1007/s10209-014-0348-1
- Marzban, A., and Jalali, F. E. (2016). The interrelationship among L1 writing skills, L2 writing skills, and L2 proficiency of Iranian EFL learners at different proficiency levels. *Theory Pract. Lang. Stud.* 6:1364. doi: 10.17507/tpls.0607.05
- Moon, J. W., and Kim, Y. G. (2001). Extending the TAM for a world-wide-web context. *Inf. Manag.* 38, 217–230. doi: 10.1016/S0378-7206(00)00061-6
- Muthén, L. K., and Muthén, B. O. (1998–2012). *Mplus User's guide (7th Edn.)*. Los Angeles, CA: Muthén & Muthén.
- Nation, I. S., and Nation, I. S. P. (2001). *Learning vocabulary in another language (10)*. Cambridge: Cambridge University Press.
- Ngo, T. T. N., Chen, H. H. J., and Lai, K. K. W. (2022). The effectiveness of automated writing evaluation in EFL/ESL writing: a three-level meta-analysis. *Interact. Learn. Environ.* 1-18, 1–18. doi: 10.1080/10494820.2022.2096642
- Nunes, A., Cordeiro, C., Limpo, T., and Castro, S. L. (2022). Effectiveness of automated writing evaluation systems in school settings: a systematic review of studies from 2000 to 2020. *J. Comput. Assist. Learn.* 38, 599–620. doi: 10.1111/jcal.12635
- Palermo, C., and Wilson, J. (2020). Implementing automated writing evaluation in different instructional contexts: a mixed-methods study. *J. Writ. Res.* 12, 63–108. doi: 10.17239/jowr-2020.12.01.04
- Parmaxi, A. (2023). Virtual reality in language learning: a systematic review and implications for research and practice. *Interact. Learn. Environ.* 31, 172–184. doi: 10.1080/10494820.2020.1765392
- Parra, G. L., and Calero, S. X. (2019). Automated writing evaluation tools in the improvement of the writing skill. *Int. J. Instr.* 12, 209–226. doi: 10.29333/iji.2019.12214a
- Polio, C. G. (1997). Measures of linguistic accuracy in second language writing research. *Lang. Learn.* 47, 101–143. doi: 10.1111/0023-8333.31997003
- Qassemzadeh, A., and Soleimani, H. (2016). The impact of feedback provision by Grammarly software and teachers on learning passive structures by Iranian EFL learners. *Theory Pract. Lang. Stud.* 6, 1884–1894. doi: 10.17507/tpls.0609.23
- Rahimi, M., and Fathi, J. (2022). Exploring the impact of wiki-mediated collaborative writing on EFL students' writing performance, writing self-regulation, and writing self-efficacy: a mixed methods study. *Comput. Assist. Lang. Learn.* 35, 2627–2674. doi: 10.1080/09588221.2021.1888753
- Ranalli, J. (2018). Automated written corrective feedback: how well can students make use of it? *Comput. Assist. Lang. Learn.* 31, 653–674. doi: 10.1080/09588221.2018.1428994
- Ranalli, J., Link, S., and Chukharev-Hudilainen, E. (2017). Automated writing evaluation for formative assessment of second language writing: investigating the accuracy and usefulness of feedback as part of argument-based validation. *Educ. Psychol.* 37, 8–25. doi: 10.1080/01443410.2015.1136407
- Roscoe, R. D., Wilson, J., Johnson, A. C., and Mayra, C. R. (2017). Presentation, expectations, and experience: sources of student perceptions of automated writing evaluation. *Comput. Hum. Behav.* 70, 207–221. doi: 10.1016/j.chb.2016.12.076
- Saricaoglu, A., and Bilki, Z. (2021). Voluntary use of automated writing evaluation by content course students. *ReCALL* 33, 265–277. doi: 10.1017/S0958344021000021
- Stapleton, P., and Radia, P. (2010). Tech-era L2 writing: towards a new kind of process. *ELT J.* 64, 175–183. doi: 10.1093/elt/ccp038
- Stevenson, M., and Phakiti, A. (2014). The effects of computer-generated feedback on the quality of writing. *Assess. Writ.* 19, 51–65. doi: 10.1016/j.asw.2013.11.007
- Stevenson, M., and Phakiti, A. (2019). Automated feedback and second language writing. *Feed. Sec. Lang. Writ.* 125–142. doi: 10.1017/9781108635547.009
- Storch, N. (2011). Collaborative writing in L2 contexts: processes, outcomes, and future directions. *Annu. Rev. Appl. Linguist.* 31, 275–288. doi: 10.1017/S0267190511000079
- Tambunan, A. R. S., Andayani, W., Sari, W. S., and Lubis, F. K. (2022). Investigating EFL students' linguistic problems using Grammarly as automated writing evaluation feedback. *Indones. J. Appl. Linguist.* 12, 16–27. doi: 10.17509/ijal.v12i1.46428
- Tang, J., and Rich, C. S. (2017). Automated writing evaluation in an EFL setting: lessons from China. *Jalt Call Journal* 13, 117–146. doi: 10.29140/jaltcall.v13n2.215
- Tsao, J. J. (2021). Effects of EFL learners' L2 writing self-efficacy on engagement with written corrective feedback. *Asia Pac. Educ. Res.* 30, 575–584. doi: 10.1007/s40299-021-00591-9
- Tuzi, F. (2004). The impact of e-feedback on the revisions of L2 writers in an academic writing course. *Comput. Compos.* 21, 217–235. doi: 10.1016/j.compcom.2004.02.003
- Uysal, H. H. (2010). A critical review of the IELTS writing test. *ELT J.* 64, 314–320. doi: 10.1093/elt/ccp026
- Venkatesh, V., and Davis, F. D. (1996). A model of the antecedents of perceived ease of use: Development and test. *Decision Sciences.* 27, 451–481. doi: 10.1111/j.1540-5915.1996.tb00860.x
- Vygotsky, L.S. (1978). *Mind in society: the development of higher psychological processes*. Cambridge, MA: Harvard University Press.
- Waer, H. (2023). The effect of integrating automated writing evaluation on EFL writing apprehension and grammatical knowledge. *Innov. Lang. Learn. Teach.* 17, 47–71. doi: 10.1080/17501229.2021.1914062
- Wallace, L. G., and Sheetz, S. D. (2014). The adoption of software measures: a technology acceptance model (TAM) perspective. *Inf. Manag.* 51, 249–259. doi: 10.1016/j.im.2013.12.003
- Wang, Y. J., Shang, H. F., and Briody, P. (2013). Exploring the impact of using automated writing evaluation in English as a foreign language university students' writing. *Comput. Assist. Lang. Learn.* 26, 234–257. doi: 10.1080/09588221.2012.655300
- Warschauer, M., and Healey, D. (1998). Computers and language learning: an overview. *Lang. Teach.* 31, 57–71. doi: 10.1017/S0261444800012970
- Warschauer, M., and Ware, P. (2006). Automated writing evaluation: defining the classroom research agenda. *Lang. Teach. Res.* 10, 157–180. doi: 10.1191/1362168806lr1900a
- Wu, B., and Chen, X. (2017). Continuance intention to use MOOCs: integrating the technology acceptance model (TAM) and task technology fit (TTF) model. *Comput. Hum. Behav.* 67, 221–232. doi: 10.1016/j.chb.2016.10.028
- Xu, J., and Zhang, S. (2022). Understanding AWE feedback and English writing of learners with different proficiency levels in an EFL classroom: a sociocultural perspective. *Asia Pac. Educ. Res.* 31, 357–367. doi: 10.1007/s40299-021-00577-7
- Youseffard, S., and Fathi, J. (2021). Exploring the impact of blogging in English classrooms: focus on the ideal writing self of EFL learners. *Int. J. Instr.* 14, 913–932. doi: 10.29333/iji.2021.14452a
- Yu, S. (2021). Feedback-giving practice for L2 writing teachers: friend or foe? *J. Second. Lang. Writ.* 52:100798. doi: 10.1016/j.jslw.2021.100798
- Zhai, N., and Ma, X. (2022). Automated writing evaluation (AWE) feedback: a systematic investigation of college students' acceptance. *Comput. Assist. Lang. Learn.* 35, 2817–2842. doi: 10.1080/09588221.2021.1897019
- Zhang, Z. V. (2020). Engaging with automated writing evaluation (AWE) feedback on L2 writing: student perceptions and revisions. *Assess. Writ.* 43:100439. doi: 10.1016/j.asw.2019.100439
- Zhang, Z. V., and Hyland, K. (2018). Student engagement with teacher and automated feedback on L2 writing. *Assess. Writ.* 36, 90–102. doi: 10.1016/j.asw.2018.02.004
- Zhang, R., Zou, D., Cheng, G., and Xie, H. (2022). Implementing technology-enhanced collaborative writing in second and foreign language learning: a review of practices, technology and challenges. *Educ. Inf. Technol.* 27, 8041–8069. doi: 10.1007/s10639-022-10941-9