

FOREIGN LANGUAGES AND VIETNAMESE

TRACKING THE COMPARATIVE EFFECTS OF TEACHER FEEDBACK VERSUS AUTOMATED FEEDBACK ON EFL/ESL WRITING: A SYSTEMATIC REVIEW OF SCOPUS-INDEXED RESEARCH

HOANG THI LINH GIANG*

ABSTRACT: Automated feedback on second language (L2) writing has increasingly gained in popularity with the advent of several automated writing evaluation (AWE) programs and artificial intelligence (AI) tools. However, research into the impact of automated feedback on L2 writing has produced inconclusive findings regarding the comparative effects of automated versus teacher feedback on students' writing performance and psychology, as well as on teachers' feedback practices. With an aim to generate significant pedagogical implications for L2 writing instruction and learning, the current study is a systematic review of relevant research studies related to the comparison between teacher and automated feedback. Adopting a systematic approach to reviewing the studies in this line of inquiry, this research analyses 19 Scopus-indexed papers, focusing on the four themes of feedback conditions, focus of comparison, the comparative effects of different feedback conditions on L2 learners' writing performance, and the effects of AWE feedback on teachers' feedback practices. Major trends emerge from the systematic review, including the facilitative effects of the combined feedback condition (either hybrid or sandwiched) on students' writing quality and accuracy gains in L2 composition. Students were also found to engage differentially with different feedback modes, but exhibited more positive perceptions of teacher feedback than automated feedback. The main implication indicates the combined use of teacher and automated feedback so that the shortcomings of one feedback source are offset by the strengths of the other.

KEY WORDS: automated feedback; teacher feedback; comparative effects; L2 writing; feedback practices.

RECEIVED: March, 17th, 2025.

EDITED: June, 20th, 2025

1. Introduction

The area of feedback research on second language writing has progressed from traditional modes to technology enhanced forms of feedback. The advent of new feedback modalities has increased learners' choices in accessing feedback from different platforms, ranging from traditional teacher written feedback to AWE programs (e.g., Grammarly, Write&Improve) and AI-generated feedback (e.g., ChatGPT, DeepSeek). As different feedback modalities exist alongside each other, variable needs of different teaching and learning contexts are satisfied. It is, however, unclear whether different feedback sources can be coordinated to bring about the most desirable effects on learners' writing skills. Besides the use of single feedback sources for reference during writing practice, the incorporation of different feedback modes has come in different shapes, including hybrid condition where two or more feedback sources are used simultaneously, separate and parallel conditions as used in experimental research with an experimental versus control/comparison group, or the sandwiched condition when one type of feedback is implemented after another. Yet, to date, previous research has produced mixed findings on the comparative effects of teacher feedback and automated feedback on second language (L2) writing (Dikli & Bleyle, 2014; Link et al., 2020; Wang & Wang, 2012; Warden, 2000; Wilson & Czik, 2016), and not many research studies have examined how learners engage with different sources of feedback (Thi et al., 2023). By examining the comparative

*Hue University, University of Foreign Languages and International Studies;
Email: htlgiang@hueuni.edu.vn

effects of different feedback sources via the lens of systematic reviews, important pedagogical implications about the use of such feedback can be generated and technical implications regarding the refinements to AWE feedback mechanisms are made possible.

2. Background of the Study

Traditional approaches to feedback generation have long been associated with teacher feedback. Yet, the emergence of technologies has transformed feedback practices to enable “a process in which learners make sense of information about their performance and use it to enhance the quality of their work or learning strategies” (Henderson et al., 2018, p.16). The automated feedback generated by automated essay evaluation (AWE) programs is rapidly becoming a popular source of feedback in writing classrooms. Automated feedback emulates to some extent the feedback provided by writing teachers. For example, *Grammarly* provides either indirect feedback (with the problematic text section underlined in red) or direct explicit feedback which includes an indication of the error type and error explanation, as well as suggested revisions (Koltovskaia, 2020). Another program, *Write & Improve*, generates three types of diagnostic feedback which range in degree of explicitness. Specifically, direct word-level feedback provides the most explicit feedback (location, explanation, and suggested correction); indirect word-level feedback is less explicit as it flags a word or phrase and provides a brief explanation; and indirect sentence level feedback uses a color scheme to indicate the quality of each sentence in terms of its accuracy and is the least explicit (Liu & Yu, 2022).

As automated feedback is gaining grounds in different educational contexts and levels, it is of great importance to have deep understanding about its impact on L2 learners' writing. Arguments against the use of AWE programs on the basis of construct validity were advanced by several scholars. For example, AWE programs are considered to be inherently limited in its conceptualization about the writing skills as they can only measure superficial features of text lengths, syntax, and vocabulary rather than gauging the higher order skills of logical thinking, idea development, or creating coherent texts (Condon, 2013; Perelman, 2014; Wilson & Roscoe, 2018). Empirical evidence reported in much of the recent research on AWE has also suggested the implausibility of leaving L2 writers to their own devices to process and engage with automated feedback to improve their writing skills. In the English as a foreign language (EFL) context in Asia, teacher feedback is generally perceived as authoritative and reliable, although teacher feedback is not always well received among students to result in improved writing skills (Han & Hyland, 2019). A combined approach where writing teachers continue to be a mediator along with the incorporation of AWE feedback has been a predominant suggestion in a lot of previous research on the impact of AWE feedback (e.g., Chen & Cheng, 2006; Hoang, 2022; Hoang & Storch, 2024). It is also well documented that different feedback sources, feedback types, and strategies of giving feedback influence students' writing in variable ways, such as their motivation, engagement, and feedback retention (Link et al., 2020; Reynolds et al., 2021; Jiang et al., 2020).

Despite a recent surge in research on AWE feedback over the last two decades since the turn of the century, what remains largely unanswered to this date is (a) to what extent and how best teacher and automated feedback can be combined and (b) how the impact of automated feedback compares to that of teacher feedback. Divergent research results have been documented on these two issues, without conclusive findings across multiple research efforts. As written feedback from various sources becomes increasingly accessible to L2 learners, it is important to investigate how best to combine these sources to enhance learning outcomes. A review of the relevant literature reveals diverse approaches to implementing teacher and automated feedback. In some studies, the two types of feedback were provided separately to different but comparable groups of learners (e.g., Mohsen & Alshahrani), while others applied one type of feedback following the other (e.g., Thi & Nikolov, 2022). In yet other cases, both feedback types were used concurrently, with each targeting specific aspects of writing (e.g., Han & Sari, 2022). This systematic review, therefore, aims to seek answers to the following research questions:

1/What feedback conditions are found in various Scopus-indexed research studies in the last five-year period from 2019-2024?; 2/What forms the focus of comparative effects of the two feedback sources: teacher feedback versus automated feedback?; 3/Of the two feedback sources, teacher feedback versus automated feedback, which is more facilitative of learners' writing development?; 4/To what extent does the use of AWE feedback influence L2 writing teacher feedback practices?

Based on the three-tiered analyses, pedagogical and practical implications are drawn and presented accordingly. Table 1 provides a summary of all the reviewed papers by each tier of analysis for this systematic review.

Table 1. A Summary of Reviewed Studies

Study	AWE program	Tier 1: Feedback conditions	Tier 2: Focus of comparison	Tier 3: Comparative effects
Mohsen & Alshahrani (2019)	MY Access	Parallel and separate, sandwiched condition: MY Access feedback only versus sandwiched condition (TF followed by MY Access feedback)	- Overall writing quality (holistic scores generated by MY Access)	Students using the sandwiched condition outscored those using the automated feedback only.
Tian & Zhou (2020)	Pigai	Sandwiched condition: Automated feedback - peer feedback - teacher feedback cycle	- Student uptake - Student engagement	- Pigai offered the largest amount of feedback in total, but the five learners incorporated the least among the three feedback sources. - Teacher feedback was perceived the most beneficial, especially at meaning-related level. Teacher feedback, however, could be indirect and thus vague for learners to take up. Direct teacher feedback was most welcomed by learners.
Hassanzadeh & Fotoohnejad (2021)	Criterion	Parallel and separate conditions: either TF or Criterion feedback	- Overall writing quality (post-test score, using Jacobs et al.'s (1981) analytic (multi-trait) scoring rubric)	The group receiving Criterion feedback over several weeks had writing scores rising significantly from the pretest to posttest.
Wang & Han (2022)	Pigaiwang	Parallel and separate conditions: either TF or Pigaiwang	- Revision quality - Overall writing quality	- Students in the teacher feedback group scored significantly higher, but

		feedback	- Students' perceptions	had significantly lower scores on the post-test of the writing proficiency than the automated feedback group. - Students in the teacher feedback class also had significantly higher ratings on perceived usefulness and perceived ease of use of the feedback than those in the automated feedback class.
Han & Sari (2022)	Criterion	- Hybrid versus Teacher feedback condition	- Overall writing quality using analytic scoring	- The combined automated-teacher feedback condition improved students' analytic writing scores as much as the students who received full teacher feedback. - The hybrid (automated-teacher feedback combined) condition outperformed the control group (full teacher feedback) in the categories of grammar and mechanics.
Link et al. (2022)	Criterion	- Hybrid versus Teacher feedback condition	- Teacher feedback practices - Student uptake - Short and long term effects on students' CAF measures	- Using Criterion alongside teacher feedback did not significantly impact the amount of higher level teacher feedback, but the teacher who did not use Criterion tended to provide a greater amount of lower level feedback than Criterion alone. - Students seemed to revise the teacher's lower level feedback more frequently than Criterion feedback.

				- Students retained their improvement in accuracy better when they had access to Criterion.
Thi & Nikolov (2022)	Grammarly	Sandwiched condition: TF, followed by Grammarly feedback, followed by combined feedback	- Revision quality - Overall writing quality - Students' perceptions	- Grammarly feedback received the highest rate of correct revision (76.2%). - The combined condition of teacher and Grammarly feedback had the highest ratio of no revision. - After a semester receiving sandwiched feedback, students gained in their writing performance.
Zhang & Hyland (2022)	Pigai	Sandwiched condition: AWE-peer-teacher feedback revision process	Student engagement: Behavioural, cognitive, and affective	- Most students actively engaged with the sandwiched approach to feedback generation: It effectively promoted students' behavioural, affective, and cognitive engagement with feedback on their writing and encouraged thoughtful revisions.
Zhang & Zhang (2022)	Pigai	Parallel and separate conditions: either TF, PF, or Pigai feedback	- Use of metacognitive strategies	- No significant differences among teacher feedback, peer feedback and automated feedback in the effect on the use of language knowledge accumulation strategies and pre-planning strategies in EFL writing. - Teacher feedback had the greatest impact on language knowledge accumulation strategies, followed by automated feedback. - Automated feedback

				had a hindrance effect on the use of monitoring strategies, whereas teacher feedback had a promotive effect on the use of all metacognitive strategies.
Ajabshir & Ebadi (2023)	Grammarly	Paralell and separate conditions: either TF focused feedback or Grammarly feedback	- Overall writing quality - Syntactic complexity, accuracy, lexical diversity, and fluency (CALF)	- Overall writing quality: No significant difference between two groups. - AWE group yielded better scores in lexical diversity and syntactic complexity, while the TF group outperformed in fluency.
Mohammadi et al. (2023)	The Virtual Writing Tutor	Paralell and separate, versus hybrid conditions: automated, teacher, and hybrid feedback	- Overall writing quality - Student perceptions	- Students receiving teacher-only and blended feedback performed better in writing. - The hybrid feedback group outperformed the others in task response, the teacher feedback group in cohesion and coherence, and the automated feedback group in lexical resource. - Most of the learners perceived all feedback types as clear, but they highly preferred the hybrid condition.
Guo & Wang (2023)	ChatGPT	Paralell and separate conditions: Teacher vs ChatGPT feedback on content, organization and language issues	- Feedback quantity and types - Teacher perceptions about collaboration with ChatGPT	- Quantity: ChatGPT produced a significantly larger amount of feedback than teachers. - Quantity: ChatGPT distributed its attention relatively equally among the three feedback foci (i.e., content, organisation, and language), while TF is content- and language

				issues-focused. - Type: ChatGPT offered more directive feedback, while the teachers used informative and query feedback more frequently. - Type: ChatGPT provided more praise when commenting on the content of student essays and more summaries than the teachers when evaluating the organisation aspect.
Fan (2023)	Grammarly	Hybrid condition: Grammarly (lower level feedback) + TF (high level feedback) condition versus TF only condition (lower level students)	- Overall writing quality	- The hybrid condition group did not significantly outperform the students from the TF-only group in syntactic and lexical complexity, accuracy, and fluency.
Ebadi et al. (2023)	Grammarly	Paralell and separate, versus hybrid conditions: Grammarly+TF versus Grammarly feedback only + TF only	- Students' use of English articles	- Grammarly significantly influenced and improved their ability to use articles, with the hybrid condition group (Grammarly + teacher feedback) having improved the most.
Koltovskaia (2023)	Grammarly	Hybrid condition: Teacher uses Grammarly feedback to supplement their feedback provision	- Impact of Grammarly use to supplement TF on TF practices	- Teachers provided feedback on both global and local aspects of writing despite using Grammarly as a supplementary source. - No division of labor as in: teacher takes care of higher-order and Grammarly takes care of lower-order concerns.

Kawashima (2023)	Grammarly	Sandwiched condition: Grammarly feedback – Teacher indirect feedback – Teacher direct feedback	- Feedback sandwiched cycle on students' overall writing quality - Student perceptions	- Although Grammarly feedback was favorably received, students appreciated the direct and indirect feedback from the teacher more highly for its appropriateness and reliability. - The largest number of participants rated the teacher's <i>direct feedback</i> as the most beneficial to their writing progress.
Taskiran & Goksel (2023)	Write & Improve	Paralell and separate conditions: either TF or Write & Improve feedback	- Academic writing quality	- Participants seemed to improve academic writing skills by taking regular feedback, and their writing scores increased slightly more when receiving feedback from teachers compared to Write & Improve feedback.
Thi et al. (2023)	Grammarly	Separate and parallel conditions: within-group design	- Feedback focus/accuracy - Student engagement - Writing accuracy	- Grammarly provided accurate feedback (44.1%), with few inaccurate flaggings while the teacher feedback is meaning-focused. - Engagement varies between the two conditions: Students tended to respond to automated feedback in a selective or focused manner; Students mostly failed to integrate the meaning-focused feedback from the teacher, demonstrated in their unsuccessful revisions and limited engagement with this feedback type. - 41.1% teacher

				feedback points and 26.8% Grammarly error flaggings resulted in correct revisions.
Haddadian (2024)	Criterion	Separate and parallel conditions (TF or Criterion feedback) and hybrid condition: between groups design	- Writing accuracy - Writing apprehension (WA)	- The hybrid condition group outperformed the two other conditions; There was no significant difference between the TF only and Criterion-only feedback conditions in writing accuracy. - The hybrid condition group had more reduction in WA than the other groups; The TF group displayed more reduction in WA compared to the Criterion feedback group.

3. Methodology

Search scope

In the initial stage, a set of criteria were decided on to keep the literature search on track. These include four screening criteria: (a) the paper presents empirical evidence, (b) the research focus is on writing in English as a second or foreign language, (c) the paper is written in English, and (d) it is indexed in Scopus database. The inclusion of Scopus-indexed research aims to make sure high-quality peer-reviewed articles are included for this systematic review.

Data collection

Search strategy. The data collection procedures began with the identification of key words for the literature search. The following key words were used across the searches in the title, keywords, or abstracts. There are three keyword sets in this literature search to reflect the focus of this systematic review:

- *First keyword:* writing
- *First set of keywords:* teacher feedback, instructor feedback, teacher comments
- *Second set of keywords:* automated writing evaluation, AWE feedback, automated feedback, automatic feedback

Using the above sets of keywords, search strings are built using the logical operators OR and AND. Specifically, OR is used between the synonyms and near synonyms within each keyword set (e.g., teacher feedback OR teacher comments) while AND is used between the (sets of) keywords. This way, within each of the abstracts, there has to be at least one term about the automated feedback and one term related to teacher feedback to make sure these two concepts are mentioned as initial screening for the inclusion of both concepts within the study. The following search string was thus generated for the literature search:

- Writing AND (Teacher feedback OR Instructor feedback OR Teacher comments) AND (automated writing evaluation OR AWE feedback OR automated feedback OR automatic feedback).

The search string was fed into the different databases most relevant to educational and applied linguistics research, including Scopus, MLA International Bibliography, British Education Index, ERIC, and MLA Directory of Periodicals. The search using this string returns a total of more than 80 papers.

Inclusion. To select the relevant studies for inclusion in this systematic review, all of the titles and abstracts of the studies above were screened in the initial round to check for duplicates in the list of journal articles produced by the different databases. In the second round, the abstracts are screened again to make sure the contents meet the four criteria above. If a study satisfied all of the selection criterion, the journal it appeared in is double-checked to confirm it is a Scopus-indexed one. In the third round, the whole paper was screened for confirmation of its fit for the systematic review.

Overall, the papers that were populated from the literature search using the search strings were downloaded, including their metadata, abstracts, keywords for checking duplicates (screening step 1) and initial exclusion of irrelevant papers (screening step 2). Papers are included if they are written in English and are peer-reviewed, research articles with empirical evidence about the comparative effects of teacher versus automated feedback. The whole text of each shortlisted article is screened again to make sure it is aligned with the focus of the four research questions related to the design of different feedback conditions (i.e., automated or teacher feedback) and the comparative effects of different feedback conditions on students' writing and teachers' feedback practices. Studies looking at the hybrid feedback mode instead of two separate feedback conditions (either teacher only or automated feedback only) is also included as they examine the combined effects for generation of pedagogical implications related to the use of both modalities in the L2 writing classroom. In addition, only articles published from 2019 to 2024 are reviewed to reflect the most recent AWE tools in the most recent five-year period. After these screening steps, the dataset finalized for analyses includes 19 articles which appear in 15 high-quality journals indexed in the Scopus database.

Data analysis

Once included for the review, the text of each paper is screened carefully to analyze issues in alignment with the research questions. The papers were read and reread several times to produce the coding list using thematic analysis, a widely used method in qualitative research for identifying, analyzing, and reporting patterns within qualitative data (Kalpokas & Hecker, 2023). Drawing on the extensive mining of the text data from the 19 selected papers, the data analyses for this systematic review were conducted using a three-tiered approach. - dựa theo quan điểm của ai? Chưa có phần background.

Tier-1 coding: Feedback conditions. Three major types of feedback conditions are coded, namely parallel and separate which contains the two modes of feedback: the teacher feedback condition or automated feedback condition; sandwiched condition which implies the use of one feedback source after the other in a cycle on the same writing task; and the hybrid condition which indicates the simultaneous use of teacher feedback and automated feedback, where each feedback source tends to focus on a set of different aspects of writing skills to supplement each other.

Tier-2 coding: The focus of comparative effects. The scanning of each paper in this round aims to unveil the aspects of writing skills (for learners) or the dimensions of teacher feedback generation practices (for writing instructors) which form the focus of each study's comparison when examining different feedback conditions. For this layer of analysis, four writing dimensions were coded: overall writing quality, writing accuracy, analytic writing quality by CALF/CAF measures, students' acquisition of target forms, student engagement and uptake of feedback, revision quality, students' use of metacognitive strategies, and student perceptions. Regarding impact on teacher feedback practices, feedback quantity and feedback types are the two coded categories.

Tier-3 coding: Identification of emerging trends from the 19 reviewed studies. Idiosyncratic findings in particular studies are also reported to highlight the variability of the research findings. The

themes identified in this part of coding align with the dimensions presented in the tier-2 coding process for more systematic and logical progression of ideas.

5. Findings

Feedback conditions in Scopus-indexed research studies in the last five-year period:

Three major feedback conditions are detected in the 19 reviewed studies: Parallel and separate feedback conditions, hybrid condition, and sandwiched condition. In the parallel and separate conditions, the feedback on student writing was either provided by the teacher or the AWE program independently, then these two feedback sources are compared for their impact on L2 writing performance. Studies employing this research design include those conducted by Mohsen and Alshahrani (2019), Hassanzadeh and Fotoohnejad (2021), Wang and Han (2022), Zhang and Zhang (2022), Ajabshir and Ebadi (2023), Guo and Wang (2023), Taskiran and Goksel (2023), Thi et al. (2023).

The hybrid condition was utilized by Han and Sari (2022), Link et al. (2022), Fan (2023), and Koltovskaia (2023). Some studies combined the two approaches above and implemented parallel, separate conditions alongside hybrid condition to compare the effects across three feedback modes (e.g., Ebadi et al., 2023; Haddadian, 2024; Mohammadi et al., 2023). In the hybrid condition, the teacher tended to provide feedback on higher order skills such as content, organisation or idea development, while the AWE programs were employed to supplement the teacher feedback via their form-focused comments on mechanics and language-related issues.

Regarding the sandwiched condition which was implemented in Tian & Zhou (2020), the order of feedback was AWE feedback - peer feedback - teacher feedback. This feedback cycle was also applied in Zhang and Hyland's (2022) research when Pigai was used as the AWE tool and Mohsen and Alshahrani's (2019) study with the use of MY Access. In Thi and Nikolov's (2022) study, however, students received the teacher feedback first, then Grammarly feedback in the second draft before they received the integrated feedback from the teacher and Grammarly in the final phase. Kawashima (2023) experimented on a different cycle which started with Grammarly feedback on students' writing, which was followed by the indirect teacher feedback on lower-level issues, and finally the teacher direct feedback on higher-level issues in student writing.

Overall, the studies were quite equally divided among the three feedback designs identified in the current research, with the first strand (i.e., parallel and separate conditions) slightly more popular when individual feedback sources are placed alongside each other to produce empirical evidence about their differential impact on different L2 writing aspects.

The focus of comparative effects of different feedback conditions:

Of the reviewed research body, nine studies investigated the impact of different feedback conditions on students' overall writing quality. However, how writing quality is operationalized varies from one study to another. For example, some studies employed holistic scores generated by the AWE program to measure changes in L2 learners' writing quality (e.g., Mohsen & Alshahrani, 2019) while others depended on post-test scores provided by trained human raters using analytic (multi-trait) or holistic scoring rubrics (e.g., Han & Sari, 2022; Hassanzadeh & Fotoohnejad, 2021; Taskiran & Goksel, 2023). Only a handful of studies looked at writing accuracy in specific target linguistic forms, such as the use of article in English (Ebadi et al., 2023) or accuracy in terms of the CALF measures (Ajabshir & Ebadi, 2023).

Student engagement with different feedback modes is another focus of comparative effects, including uptake and revision (behavioral engagement, as in Link et al., 2022; Thi & Nikolov, 2022; Thi et al., 2023; Tian & Zhou, 2020; Zhang & Hyland, 2022) and student perceptions (affective engagement, as in Kawashima, 2023; Mohammadi et al., 2023; Thi & Nikolov, 2022; Wang & Han, 2022). Zhang and Zhang (2022) was the only study that looked into students' use of metacognitive strategies, touching on the cognitive dimension of the tripartite framework of student engagement with feedback. More recently, Haddadian (2024) included the psychological aspect of writing

apprehension when comparing the impact of different feedback modes. Adding to this strand of research, some studies examined the impact of AWE feedback on teacher feedback practices (e.g., Guo & Wang, 2023; Koltovskaia, 2023; Link et al., 2022). Approaching feedback comparative impact from a different perspective, two studies (i.e., Guo & Wang, 2023; Thi et al., 2023) compared automated and teacher feedback accuracy, feedback focus as well as feedback types. These two studies shed light on the performance of each feedback source which may influence student engagement and use.

Overall, previous research over the last five years have investigated the impact of teacher and automated feedback from both system-centric and user-centric perspectives. System-centric studies operationalised the effects of different feedback conditions via students' writing quality either holistically or analytically using different measures of accuracy, complexity, and fluency. In addition, such research also explored student engagement with feedback via the tripartite framework of cognitive, behavioural, and affective dimensions. The number of studies related to this area of research mushroomed in the two years of 2022 and 2023, and the largest number of studies examined Grammarly feedback (seven studies), followed by Pigai, an AWE program developed in China and Criterion developed by ETS (four studies each). The remaining studies employed ChatGPT, The Virtual Writing Tutor, MY Access, and Write&Improve, with one study for each tool.

Comparative effects of different feedback conditions on students' writing skills:

Some major trends emerged from the review of this body of research. Most noticeably, several studies employing hybrid or sandwiched conditions suggest that the combined use of teacher and automated feedback results in better writing quality among L2 learners (e.g., Ebadi et al., 2023; Han & Sari, 2022; Mohsen & Alshahrani, 2019), except for the case of Grammarly in Thi and Nikolov's (2022) enquiry when the combined feedback resulted in the largest number of no revisions among the learners and in Fan (2023) where the combined condition of Grammarly lower-order feedback and higher-order teacher feedback did not lead to better performance in terms of the CAF measures compared to the teacher-only feedback condition. When examined in parallel and separate conditions, automated feedback seems to outperform teacher feedback in terms of students' gains in writing accuracy (e.g., Hassanzadeh & Fotoohnejad, 2021; Link et al., 2022; Wang & Han, 2022). Some studies report differential effects of the teacher feedback and AWE feedback on different aspects of writing. For example, Ajabshir and Ebadi's (2023) research found no significant difference in the impact of AWE feedback and teacher feedback on overall writing quality, although varying effects on lexical diversity, syntactic complexity, and fluency were detected. Similar findings were reported by Mohammadi et al. (2023) on the performance of The Virtual Writing Tutor.

When used as a sole reference for revision, AWE feedback tends to be adopted in a selective manner (e.g., Thi et al., 2023). Compared to the trust students perceived in teacher feedback (e.g., Kawashima, 2023; Tian & Zhou, 2020; Wang & Han, 2022), L2 writers were more likely to exercise caution when processing automated feedback (Link et al., 2022). In addition, AWE programs were generally found to generate more feedback on lower level issues in student texts, which resulted in more correct revisions (e.g., Thi & Nikolov, 2022; Tian & Zhou, 2020). This, however, was not the case for Write&Improve feedback in Taskiran and Goksel's (2023) research which found slightly higher impact of teacher feedback on students' writing scores. In a similar vein, no difference in the effects of either teacher feedback or automated feedback on students' writing accuracy was found for the case of Criterion in Haddadian's (2024) recent study.

Another dimension of the impact of AWE feedback is student engagement. A hybrid condition is reported to enhance students' behavioural, cognitive, and affective engagement (e.g., Zhang & Hyland, 2022). Depending on the feedback types, students exhibited variable engagement patterns. While engaging in a limited manner with the form-focused feedback from Grammarly and Pigai (Thi et al., 2023; Tian & Zhou, 2020), students also had limited engagement with the meaning-focused feedback from the teacher (Thi et al., 2023). Taking a focused perspective on students' use of

metacognitive strategies, Zhang and Zhang (2022) found teacher feedback to have promotive effects on students' planning (including language knowledge accumulation strategies and pre-planning strategies), monitoring (including selective attention strategies and self-monitoring strategies), and self-evaluation strategies. The greatest impact of teacher feedback was on students' knowledge accumulation strategies. On the contrary, the automated feedback generated by Pigai was found to hinder the use of monitoring strategies.

Comparative effects of different feedback conditions on teachers' feedback practices:

On the impact of AWE programs on teacher feedback practices, the use of Criterion did not result in changes to the amount of teacher feedback on higher level issues such as coherence and organisation (Criterion in Link et al., 2023; Grammarly in Koltovskaia, 2023). The few studies in this strand of research has generally pointed out that the writing teachers did not tend to rely on AWE programs to replace their form-focused feedback. These findings are contrary to the widely held belief that while AWE can address lower-order concerns, writing teachers should devote more time and efforts to delivering meaning-focused feedback or feedback on higher-order issues in students' compositions. ChatGPT, an AI-driven feedback mechanism, provides a different perspective on how automated feedback can be incorporated in L2 writing. In Guo and Wang's (2023) research on five Chinese EFL teachers, ChatGPT was found to generate a much larger amount of feedback which was equally distributed across three different dimensions of content, organisation, and language, while the teacher feedback was focused on meaning and language issues. The language to communicate the feedback to students was also found to be different, with ChatGPT delivering more directive and praising comments, and writing teachers offered informative and query feedback besides the appraisal of organisation in students' essays.

6. Discussion and Conclusions

The systematic review of recent research on the comparative effects of automated feedback versus teacher feedback has showed some divergent findings. Such differences could be attributed to the implementation of different AWE programs in the reviewed studies. For example, the nature of ChatGPT feedback which is dependent on the input prompts devised by human users is different from the feedback automatically generated by other AWE programs such as Criterion or Grammarly. In addition, various AWE platforms have different feedback mechanisms. While Grammarly offers direct corrective feedback with clear suggested corrections, Write&Improve includes both direct and indirect word-level feedback, besides the indirect sentence-level feedback in the form of highlighted segments in the compositions. Taking these differences into account, it is recommended that future research studies include the moderating factors of the types of feedback in their measure of the comparative effects on L2 writing and feedback practices.

Another interesting finding is the fact that writing instructors do not depend on AWE feedback to reduce their workload in generating feedback on L2 writing. In the studies which integrate AWE feedback and teacher feedback, the teachers still provided feedback on both lower-level and higher-order skills in L2 students' writing. In other words, the use of AWE feedback adds to rather than replacing teacher feedback provision. This result finds additional support from the empirical evidence in most reviewed studies that the combined condition, whether it be the hybrid or sandwiched mode, is facilitative of overall writing quality or accuracy. Drawing on the positive perceptions of teacher feedback while mining the technical capacities of automated feedback is a promising strategy to get the best of both worlds, as pointed out by Wang and Han (2022). With its instant feedback mechanism and the large amount of feedback, AWE programs can improve writing instructors' feedback generation in different ways. Firstly, writing teachers can speed up their provision of feedback by using AWE programs to generate and refine the automated feedback reports or implement AI tools to devise relevant focused feedback that can be communicated to learners (Hassanzadeh & Fotoohnejad, 2021). Secondly, the use of sandwiched feedback conditions can be adapted to suit specific learner groups and teaching contexts. Depending on learner proficiency levels

and needs, the feedback cycle can adopt the form-then-content sequence, leaving form-focused feedback from AWE programs to do its job before teacher feedback comes into play (Ebadi et al., 2023; Haddadian, 2024; Kawashima, 2023; Guo & Wang, 2023; Tian & Zhou, 2020). This proposed integrated approach to feedback generation is empirically founded on the recent research in this line of inquiry which tends to agree on the supplementary roles of teacher and automated feedback. More specifically, the instantaneous automated feedback makes up for the delayed nature of teacher feedback, while the teacher feedback can address the lack of graduated feedback from AWE platforms.

Acknowledgements: *This research was supported by Hue University's Research Grant under Grant No. DHH2025-07-112.*

REFERENCES

1. Chen, C. E., & Cheng, W. (2006), *The use of computer-based writing program: Facilitation or frustration?* Retrieved from ERIC database. (ED492078).
2. Dikli, S., & Bleyle, S. (2014), Automated essay scoring feedback for second language writers: How does it compare to instructor feedback. *Assessing Writing*, 22, 1-17. <https://doi.org/10.1016/j.asw.2014.03.006>.
3. Hoang, T. L. G. (2022), Feedback precision and learners' response: A study into ETS Criterion automated corrective feedback in EFL writing classrooms. *The JALT CALL Journal*, 18(3), 444-467.
4. Henderson, M., Phillips, M., Ryan, T., Boud, D., & Dawson, P. (2018). Conditions that enable effective feedback. *Higher Education Research & Development*, 38(7), 1-16. <https://doi.org/10.1080/07294360.2019.1657807>.
5. Hoang, T. L. G., & Storch, N. (2024), How do EFL learners process and uptake Criterion automated corrective feedback? Insights from two case studies. *Iranian Journal of Language Teaching Research*, 12(1), 1-20. <http://ijltr.urmia.ac.ir>.
6. Jiang, L., Yu, S., & Wang, C. (2020), Second language writing instructors' feedback practice in response to automated writing evaluation: A sociocultural perspective. *System*, 93. <https://doi.org/10.1016/j.system.2020.102302>.
7. Kalpokas, N., & Hecker, J. (2023), The ultimate guide to qualitative research - Part 3. ATLAS.ti Research Hub. <https://atlasti.com/guides/qualitative-research-guide-part-3>
8. Koltovskaia, S. (2020). Student engagement with automated written corrective feedback (AWCF) provided by Grammarly: A multiple case study. *Assessing Writing*, 44. <https://doi.org/10.1016/j.asw.2020.100450>.
9. Link, S., Mehrzad, M., & Rahimi, M. (2020), Impact of automated writing evaluation on teacher feedback, student revision, and writing improvement. *Computer Assisted. Language Learning*, 35(4), 605-634. <https://doi.org/10.1080/09588221.2020.1743323>.
10. Liu, S., & Yu, G. (2022), L2 learners' engagement with automated feedback: An eye-tracking study. *Language Learning & Technology*, 26(2), 78-105. <https://doi.org/10.125/73480>.
11. Reynolds, B. L., Kao, C. W., & Huang, Y. (2021), Investigating the effects of perceived feedback source on second language writing performance: A quasi-experimental study. *Asia-Pacific Education Researcher* 30(6), 585-595. <http://dx.doi.org/10.1007/s40299-021-00597-3>.
12. Thi, N. K., Nikolov, M., & Simon, K. (2023), Higher-proficiency students' engagement with and uptake of teacher and Grammarly feedback in an EFL writing course. *Innovation in Language Learning and Teaching*, 17(3), 690-705. <https://doi.org/10.1080/17501229.2022.2122476>.

13. Wang, F., & Wang, S. (2012), A comparative study on the influence of automated evaluation system and teacher grading on students' English writing. *Procedia Engineering*, 29, 993-997. <https://doi.org/10.1016/j.proeng.2012.01.077>.
14. Warden, C. A. (2000), EFL business writing behaviors in differing feedback environments. *Language Learning*, 50(4), 573-616. <https://doi.org/10.1111/0023-8333.00141>.
15. Wilson, J., & Czik, A. (2016), Automated essay evaluation software in English Language Arts classrooms: Effects on teacher feedback, student motivation, and writing quality. *Computers & Education*, 100, 94-109. <https://doi.org/10.1016/j.compedu.2016.05.004>.

So sánh tác động của phản hồi từ giáo viên và phản hồi tự động đối với kỹ năng viết tiếng Anh: Đánh giá hệ thống dựa trên các nghiên cứu trong cơ sở dữ liệu Scopus

Tóm tắt: Phản hồi tự động đối với kỹ năng viết tiếng Anh như một ngôn ngữ thứ hai ngày càng trở nên phổ biến với sự ra đời của các chương trình đánh giá viết tự động (AWE) và các công cụ trí tuệ nhân tạo (AI). Tuy nhiên, nghiên cứu về ảnh hưởng của phản hồi tự động đối với kỹ năng viết đã cho ra những kết quả thiếu nhất quán, và do đó còn thiếu bằng chứng thực nghiệm về mức độ tác động của phản hồi tự động so với phản hồi từ giáo viên đối với kỹ năng viết và tâm lý của người học, cũng như đối với hoạt động phản hồi bài viết của giáo viên. Với mục đích nhắm tới những khuyến nghị có ý nghĩa cho quá trình giảng dạy và học kỹ năng viết tiếng Anh như một ngoại ngữ, bài viết này là một đánh giá có hệ thống về các nghiên cứu liên quan đến sự so sánh giữa phản hồi của giáo viên và phản hồi tự động. Áp dụng phương pháp tổng quan hệ thống để xem xét các nghiên cứu trong lĩnh vực này, bài báo phân tích 19 nghiên cứu thực nghiệm nằm trong cơ sở dữ liệu Scopus, tập trung vào bốn chủ đề: cách thức thực hiện phản hồi, trọng tâm tác động được so sánh, mức độ tác động của các cách thức tiến hành phản hồi khác nhau đối với kỹ năng viết của người học, và tác động của phản hồi tự động đối với quá trình đưa ra phản hồi bài viết của giáo viên. Những xu hướng chính được phát hiện bao gồm: sự kết hợp hai hình thức phản hồi của giáo viên và phản hồi tự động (theo cách kết hợp song song hoặc xen kẽ) có tác động tích cực nhất đối với chất lượng và độ chính xác trong bài viết tiếng Anh của người học. Bài đánh giá hệ thống cũng cho thấy người học có cách tương tác khác nhau với các hình thức phản hồi khác nhau, nhưng nhìn chung họ đánh giá phản hồi của giáo viên cao hơn so với phản hồi tự động. Từ đó, bài báo này khuyến nghị việc sử dụng kết hợp phản hồi của giáo viên và phản hồi tự động để hai nguồn phản hồi có thể được bù đắp cho nhau, mang lại hiệu quả tốt nhất cho quá trình dạy và học kỹ năng viết trong tiếng Anh.

Từ khóa: phản hồi tự động; phản hồi của giáo viên; mức độ tác động; viết ngôn ngữ thứ hai; hoạt động phản hồi bài viết.