

Effects of AI-Generated Direct and Metalinguistic Written Corrective Feedback on Grammatical Accuracy: Evidence from the English Passive Causative

Nargess Nourizadeh¹  Amir Valadi²  Sayyed Behnam Alavi Moghaddam³ 

¹ Assistant Professor, Department of English Language and Literature, Faculty of Humanities, Lorestan University, Khorramabad, Iran. (Corresponding Author)
E-mail: nourizadeh.n@lu.ac.ir

² Assistant Professor, Department of Persian Language and Literature Education, Farhangian University, P. O. Box 14665-889, Tehran, Iran. E-mail: amirvaladi2020@cfu.ac.ir

³ Associate Professor, Department of Foreign Languages, Educational Research and Planning Organization, Tehran, Iran. E-mail: behnamalavim@gmail.com

Article Info

Article type:

Research article

Article history:

Received: 1 Jan. 2026

Received in revised form:
11 Sep. 2026

Accepted: 15 Oct. 2026

Published online:

Keywords:

written Corrective Feedback,
direct feedback,
metalinguistic feedback,
artificial Intelligence,
passive causative

ABSTRACT

The growing use of artificial intelligence (AI) in second language writing has increased interest in AI-generated written corrective feedback; however, little is known about how different feedback types influence the retention of grammatical knowledge over time. Grounded in form-focused instruction and written corrective feedback research, this study examined the effects of AI-generated corrective feedback and metalinguistic feedback on EFL learners' sentence-level grammatical accuracy. Fifty-eight university students were assigned to two feedback conditions, and their accuracy was assessed through a pretest, immediate posttest, and delayed posttest targeting the English passive causative construction. Both feedback types produced substantial improvements in grammatical accuracy on the immediate posttest. However, learners who received AI-generated metalinguistic feedback maintained higher accuracy on the delayed posttest than those receiving direct corrective feedback, indicating different developmental trajectories. These findings highlight the importance of feedback explicitness in AI-mediated grammar instruction and offer pedagogical implications for integrating AI into writing instruction.

Cite this article: Nourizadeh, N., Valadi, A., & Alavi Moghaddam, S. B. (2026). Effects of AI-Generated Direct and Metalinguistic Written Corrective Feedback on Grammatical Accuracy: Evidence from the English Passive Causative. *Journal of Linguistic Studies: Theory and Practice*, Year (Vol.), ...-.....



© The Author(s).

Publisher: University of Kurdistan.

DOI: [10.22034/jls.2026.145278.1339](https://doi.org/10.22034/jls.2026.145278.1339)

Corresponding author: nourizadeh.n@lu.ac.ir

Introduction

Providing effective written corrective feedback (WCF) remains one of the most demanding aspects of second language (L2) writing instruction. In many EFL classrooms, teachers are expected to provide timely and individualized feedback on learners' grammatical errors despite constraints on instructional time and workload. Recent advances in artificial intelligence (AI), particularly the emergence of large language models such as ChatGPT, have introduced new possibilities for supporting feedback provision by generating context-sensitive responses that can be delivered consistently and efficiently (Teng, 2024). These developments have renewed interest in how AI-generated feedback may contribute to learners' grammatical development in L2 writing.

Within the broader literature on written corrective feedback, considerable attention has been devoted to feedback explicitness, particularly the distinction between direct corrective feedback (DCF) and metalinguistic feedback (MF). DCF provides learners with the correct linguistic form, whereas MF encourages learners to reflect on the underlying grammatical rule (Ellis, 2009). Previous research suggests that these feedback types may lead to different learning outcomes, although findings remain mixed across learner populations, instructional contexts, and target structures (Bitchener & Knoch, 2010; Li, 2010; Shintani & Ellis, 2013; Suzuki et al., 2019).

Building on these established distinctions in written corrective feedback, generative AI now provides a new means of delivering different forms of corrective feedback in language teaching contexts. Unlike earlier automated writing evaluation (AWE) systems, which were largely limited to surface-level error detection and formulaic feedback (Ranalli et al., 2017; Saricaoglu, 2019), generative AI tools are capable of producing extended, context-sensitive responses, including grammatical explanations, reformulations, and adaptive feedback tailored to learner input (Kohnke et al., 2023). These affordances have generated considerable interest in AI as a potential means of addressing long-standing challenges in feedback provision, such as large class size, limited classroom time, and unequal access to individualized support. Consistent with this growing interest, emerging evidence suggests that AI-generated feedback can support learners' revision practices, engagement, and self-efficacy and may approximate human feedback in terms of clarity and elaboration (Eliebyary & Shabara, 2024; Yang et al., 2025).

Despite these developments, much of the existing research on AI-mediated WCF has focused on learners' perceptions, revision behaviors, or holistic writing quality (e.g., Zhang & Hyland, 2018; Yan & Zhang, 2024; Cengiz et al., 2025), rather than on structure-specific grammatical development measured through controlled sentence-level production. While these studies provide valuable insights into learners' engagement with AI-generated feedback, they offer limited evidence regarding whether such feedback supports the maintenance of grammatical accuracy over time. Moreover, few studies have directly compared different AI-generated feedback types under controlled experimental conditions using delayed measures of grammatical development.

Research on form-focused instruction further suggests that the effectiveness of feedback is influenced by multiple interacting variables, including feedback explicitness, the nature and intensity of instruction,

and the characteristics of the target structure (Ellis, 2002). Ellis (2002) notes that instructional success cannot be explained solely by structural complexity or the amount of instruction, as studies have reported positive learning outcomes under a range of instructional conditions. At the same time, learners may vary in the extent to which they transform explicit grammatical knowledge into consistent and spontaneous use over time (DeKeyser, 2007). These considerations raise important questions in the context of AI-mediated feedback, particularly regarding whether different types of AI-generated written corrective feedback support grammatical development in similar or divergent ways across time.

In the current research, the English passive causative construction (e.g., *I had my phone repaired*) was selected as the target structure because it presents difficulty for many EFL learners due to its complex form–meaning relationships and the non-agentive role of the grammatical subject. As a relatively infrequent and structurally demanding feature, the passive causative provides an appropriate context for examining whether different forms of AI-generated feedback support learners' accuracy development over time.

Accordingly, the present study investigates the effects of AI-generated direct written corrective feedback and AI-generated metalinguistic feedback on EFL learners' grammatical accuracy across three testing occasions: pretest, immediate posttest, and delayed posttest. Using a 2×3 mixed-design quasi-experimental approach, the study explores whether feedback type and time interact to shape learners' grammatical accuracy in producing the English passive causative construction. By incorporating a delayed posttest and focusing on a specific grammatical structure, the study moves beyond immediate revision effects and contributes to ongoing debates on feedback explicitness, the maintenance of grammatical gains, and the pedagogical role of generative AI in L2 writing instruction.

A brief note of previous works

Written corrective feedback and feedback explicitness

Research on written corrective feedback has consistently shown that appropriately designed feedback can facilitate second language development by directing learners' attention to linguistic form (Ellis, 2002; Lyster & Ranta, 1997). Although early critiques questioned the pedagogical value of grammar correction in writing instruction, most notably Truscott (1996), subsequent empirical and meta-analytic research has provided strong evidence that written corrective feedback can contribute to grammatical development when factors such as learner characteristics, proficiency level, instructional context, and feedback design are taken into account (Bitchener & Knoch, 2010; Li, 2010; Mao et al., 2024).

A well-established distinction in the written corrective feedback literature concerns direct corrective feedback and metalinguistic feedback (Ellis, 2009). Direct corrective feedback supplies learners with the correct linguistic form, which can ease processing demands and support rapid error correction. Metalinguistic feedback, by contrast, provides learners with grammatical explanations, rules, or error codes that require reflection on the underlying linguistic system. One explanation for the potential effectiveness of metalinguistic feedback comes from Skill Acquisition Theory (DeKeyser, 2007), which

proposes that explicit grammatical knowledge can gradually become proceduralized through practice. From this perspective, metalinguistic feedback may facilitate learners' reflection on grammatical rules and support subsequent grammatical performance, particularly for complex structures.

Empirical studies comparing direct corrective feedback and metalinguistic feedback have yielded mixed but informative findings. Rather than demonstrating the consistent superiority of one feedback type, previous research suggests that their relative effectiveness depends on factors such as the target structure, instructional conditions, and the outcomes used to measure learning. For example, Shintani et al. (2014) found that direct corrective feedback followed by revision produced the most durable improvement for a complex syntactic structure, whereas Rezazadeh et al. (2015) reported that metalinguistic explanation resulted in better maintenance of both explicit and implicit knowledge of English articles over time. Similarly, Shintani and Ellis (2013) showed that the effects of written corrective feedback varied according to the type of knowledge assessed and the grammatical feature under investigation. Collectively, these studies suggest that the effectiveness of written corrective feedback cannot be generalized independently of the linguistic target and instructional context.

Importantly, the effectiveness of written corrective feedback should not be evaluated solely on the basis of immediate performance. Li's (2010) meta-analysis highlights the value of delayed posttests for distinguishing short-term performance gains from learning that is maintained over time. This consideration is particularly relevant when feedback targets relatively complex grammatical structures. Reviewing research on form-focused instruction, Ellis (2002) argues that instructional effectiveness depends on multiple interacting factors, including the characteristics of the target structure and the intensity of instruction. Although complex structures often benefit less from brief instructional treatments than simpler features, Ellis (2002) also notes that successful learning is possible under a range of instructional conditions. These observations make it important to examine how different forms of AI-generated written corrective feedback influence learners' development over time when instruction focuses on a relatively complex grammatical structure.

Automated and AI-mediated written corrective feedback

Alongside traditional research on written corrective feedback, an expanding body of work has examined automated approaches to feedback in second language writing. Early automated writing evaluation systems were shown to increase the availability of feedback to learners and support learner autonomy. Nevertheless, concerns remained regarding the accuracy of automated feedback across different error types and learners' ability to interpret and use that feedback effectively (Ranalli et al., 2017). Subsequent research has also emphasized that the effectiveness of automated written corrective feedback varies across learner characteristics and instructional contexts (Shiyao, 2021). Related research on sustained assessment practices has also highlighted the importance of learners' continued engagement with evaluative information. Rahmani et al. (2025), for example, found that the ongoing use of structured peer- and self-assessment was associated with improvement in EFL learners' writing over a semester. Although their study did not examine automated corrective feedback, it underscores the importance of repeated

engagement with feedback and assessment rather than viewing evaluative information as a one-time intervention. Comparative studies further suggest that automated and teacher feedback each possess distinct strengths, with their effectiveness influenced by factors such as learner engagement, writing task, and the linguistic features being assessed (Ajabshir & Ebadi, 2023; Zhang & Hyland, 2018).

Recent technological developments, particularly the emergence of generative artificial intelligence tools such as ChatGPT, have renewed scholarly interest in automated written corrective feedback. Unlike earlier automated writing evaluation systems, generative AI is capable of producing extended, adaptive, and linguistically explicit feedback that supports interactive revision processes. Recent studies suggest that learners actively engage with ChatGPT-generated feedback during revision, although the nature and extent of this engagement vary across feedback types and individual learners (Cengiz et al., 2025; Yan & Zhang, 2024). Learners have also generally perceived AI-generated feedback as helpful, detailed, and supportive, while also reporting challenges related to feedback length, specificity, and interpretation (Cengiz et al., 2025; ElEbyary & Shabara, 2024; Yan & Zhang, 2024).

Comparative research has also begun to examine AI-generated feedback alongside teacher feedback. Muñoz et al. (2025) found that ChatGPT-generated direct corrective feedback produced grammatical range and accuracy gains comparable to those achieved through teacher feedback, and suggested integration of AI tools to improve corrective feedback provision. Similarly, Zhang (2024) reported that the use of ChatGPT as a feedback tool enhanced learner motivation and revision quality, while also emphasizing the importance of pedagogical guidance to prevent superficial reliance on AI output. This broader point is also reflected in research on technology-mediated language learning. Mehrara et al. (2025), for example, found that guided mobile-assisted task performance was particularly beneficial for learners' willingness to communicate and reactivity to mistakes, suggesting that the pedagogical guidance accompanying a technological tool can influence how productively its affordances are realized. Although their study did not concern generative AI or written corrective feedback, its findings reinforce the view that technology-mediated learning outcomes cannot necessarily be attributed to the technological tool alone. More broadly, Urzúa et al. (2025), in a systematic review of AI-assisted feedback in higher education, concluded that generative AI has considerable potential to improve writing accuracy, learner autonomy, and motivation, while emphasizing that its effectiveness depends on informed pedagogical implementation and appropriate teacher mediation.

Recent analyses also suggest that generative AI differs from teacher feedback not only in its immediacy but also in its scope. For example, Guo and Wang (2024) found that ChatGPT generated substantially more feedback than teachers and distributed its attention more evenly across content, organization, and language, whereas teacher feedback focused primarily on content- and language-related issues. These findings illustrate the flexibility of generative AI as a feedback provider, although they also suggest that the nature of AI-generated feedback may differ considerably depending on the instructional prompts and feedback objectives.

Despite these advances, much of the existing research on AI-mediated written corrective feedback has focused on learner perceptions, revision behavior, or overall writing quality rather than on structure-specific grammatical development across time. Moreover, relatively few studies have examined how different types of AI-generated feedback, such as direct and metalinguistic feedback, shape learners' developmental trajectories. Given evidence from traditional feedback research showing that feedback explicitness interacts with grammatical complexity and retention (Suzuki et al., 2019), there is a clear need for experimental studies that combine AI-mediated feedback with delayed measures of grammatical development.

To address this gap, the present study compared the effects of AI-generated direct corrective feedback and AI-generated metalinguistic feedback on learners' accuracy in producing the English passive causative construction across pretest, immediate posttest, and delayed posttest measures. Based on this gap, the following research questions guided the study:

RQ1. Does EFL learners' sentence-level accuracy in producing the English passive causative construction change significantly across the pretest, immediate posttest, and delayed posttest following AI-generated written corrective feedback?

RQ2. Does the development of sentence-level grammatical accuracy over time differ between learners receiving AI-generated metalinguistic feedback and those receiving AI-generated direct corrective feedback?

RQ3. Are gains in sentence-level grammatical accuracy better maintained following AI-generated metalinguistic feedback than following AI-generated direct corrective feedback, as measured by the delayed posttest?

Theoretical framework

This study is informed by the principles of form-focused instruction (FFI), which emphasizes integrating attention to linguistic form within meaning-oriented language use (R. Ellis, 2015). Within this perspective, written corrective feedback directs learners' attention to discrepancies between their own production and target language forms, thereby creating opportunities for noticing relevant grammatical features (Ellis, 2009). Research on form-focused instruction has generally shown that explicit instructional techniques can promote second language development and that these effects may be maintained over time. One explanation for these effects is that explicit instruction helps learners establish form–function connections that can subsequently be strengthened through continued language use (N. Ellis, 2015).

Feedback explicitness represents an important dimension of FFI. Both direct corrective feedback and metalinguistic feedback are generally regarded as explicit forms of feedback, although they differ in the type of information they provide to learners (Ellis, 2009). Direct corrective feedback supplies learners with the correct linguistic form, whereas metalinguistic feedback encourages reflection on the grammatical rule underlying the error. Previous research suggests that the effectiveness of these feedback

types may vary depending on learner characteristics, instructional context, and the complexity of the target structure (Ellis, 2002; Ellis, 2009; Shintani & Ellis, 2013).

In AI-mediated language instruction, generative AI provides a new means of delivering established forms of corrective feedback rather than introducing fundamentally new learning mechanisms. Against this theoretical background, the present study examines whether AI-generated direct corrective feedback and AI-generated metalinguistic feedback are associated with different patterns of grammatical development over time.

Method

Research design

This study adopted a quasi-experimental design with a pretest, an immediate posttest, and a delayed posttest to examine the effects of two types of AI-generated written corrective feedback on EFL learners' grammatical accuracy. Specifically, the study compared AI-generated metalinguistic feedback with AI-generated direct corrective feedback in learners' sentence-level production of the English passive causative construction. The inclusion of a delayed posttest made it possible to examine not only immediate learning gains, but also the extent to which any observed effects were sustained over time.

Participants

Participants were 58 first-year undergraduate students enrolled in a compulsory General English course at Farhangian University in Khorramabad. Because all participants were enrolled in the same course and had completed the same national English curriculum prior to university, they represented a relatively homogeneous educational background. Although no standardized proficiency test was administered, participants had completed the same national English curriculum prior to university. During the course orientation session, all participants reported that they had not received any additional formal English instruction beyond the national curriculum.

The study was conducted using two intact classes, each of which was assigned to one feedback condition. One class received AI-generated metalinguistic feedback, while the other received AI-generated direct corrective feedback. Participation in the study was voluntary, and informed consent was obtained from all students prior to data collection. Participants were explicitly informed that their decision to participate or withdraw would have no effect on their course grades or academic standing.

Target structure

The instructional focus of the study was the English passive causative construction, operationalized as have or get followed by an object and a past participle, as in '*I had my phone repaired*'. This structure was selected because it is a rule-governed grammatical construction involving a complex form–meaning relationship, particularly the non-agentive role of the grammatical subject. These characteristics make it an appropriate target for examining learners' grammatical accuracy and the potential effects of different feedback types. To ensure structural consistency and scoring clarity, the study focused exclusively on passive causative forms, and active causative constructions were excluded from both instructional and assessment materials. Because this structure is not introduced in Iranian high school English textbooks

and is typically not covered in introductory university General English courses, and participants had not received instruction beyond the elementary levels of university English courses, prior exposure to the target structure was unlikely. This assumption was supported by the low accuracy scores observed on the pretest.

Instruments

Learners' grammatical accuracy was assessed using a pretest, an immediate posttest, and a delayed posttest administered three weeks after the posttest. Each test was designed to elicit sentence-level production of the passive causative construction and included two task types, sentence completion and guided sentence writing. The same test format was used across testing occasions with minor lexical modifications to selected content words in order to reduce test–retest effects while maintaining the same grammatical focus and task requirements.

The sentence completion task consisted of ten items requiring learners to supply an appropriate passive causative form based on contextual cues. The guided sentence writing task included five prompts that required learners to produce a complete sentence using the target structure in a specified context. All three test versions were reviewed by two ELT experts to ensure that they were comparable in terms of grammatical targets, task format, level of difficulty, and linguistic appropriateness for the participants. The parallel forms differed only in contextual content while The parallel forms differed only in contextual content, while preserving equivalent grammatical targets and response demands.

Treatment procedure

At the beginning of the study, both groups completed the pretest during regular class time. No feedback was provided on the pretest in order to capture learners' baseline ability to produce the target structure. Given that the passive causative had not been explicitly taught in any prior course taken by the participants, pretest performance was expected to be low and comparable across groups.

Following the pretest, both groups received a brief teacher-led instructional session, lasting approximately 15 minutes, introducing the English passive causative construction, including its grammatical form, meaning, and common communicative uses. The instructional content was identical for both groups. Learners then completed a treatment activity consisting of eight sentence-production tasks designed to elicit the passive causative construction. After submitting each response, learners received AI-generated written corrective feedback according to their assigned condition.

In the AI-generated direct corrective feedback condition, ChatGPT was guided by a fixed researcher-designed prompting protocol that instructed the system to provide only the corrected version of an erroneous sentence without grammatical explanation or translation, and then ask learners to rewrite the sentence using the corrected form. In the AI-generated metalinguistic feedback condition, ChatGPT was guided by a parallel prompting protocol instructing the system to identify errors related to the passive causative construction, explain the relevant grammatical rule in Persian, refrain from providing the corrected sentence directly, and ask learners to rewrite their sentences independently. Persian was used for the metalinguistic explanations to ensure that learners clearly understood the grammatical rule being

explained, allowing the two treatment conditions to differ primarily in the type of feedback provided rather than in learners' ability to comprehend the feedback.

In both conditions, learners were required to revise their original sentences after receiving feedback. To ensure consistency across participants, learners did not interact freely with ChatGPT. Instead, all feedback was generated using standardized researcher-designed prompts under the instructor's supervision, who monitored the session to ensure that the feedback remained focused on the target structure and that all learners completed the required revisions.

The immediate posttest was administered during the same week as the treatment session. The delayed posttest was administered three weeks later to examine the extent to which gains in grammatical accuracy were maintained over time. A three-week interval was selected because it is commonly used in written corrective feedback research to distinguish immediate learning gains from short-term retention while minimizing the influence of additional instruction (Rezazadeh et al., 2015). All test responses were scored dichotomously. One point was awarded for each grammatically accurate passive causative sentence, and zero points were assigned to inaccurate responses. Inaccurate responses included incorrect tense marking on have or get, incorrect past participle forms, incorrect word order, and the use of simple active or passive constructions that failed to convey causative meaning. Raw scores ranging from 0 to 15 were calculated for each test and used directly in all statistical analyses.

Scoring procedure

To establish scoring reliability, a second trained rater independently scored approximately twenty-five percent of the test data, randomly sampled across both feedback conditions and all testing occasions. Inter-rater agreement was calculated using Cohen's kappa and indicated substantial agreement between raters, with a kappa value of .83. Any discrepancies were resolved through discussion prior to final scoring.

Data analysis

Before statistical analysis, the dataset was carefully screened for completeness and accuracy. All participants completed the pretest, posttest, and delayed posttest, and no missing data were identified. Potential outliers were examined using standardized z-scores for test performance at each testing occasion. As no scores exceeded plus or minus three standard deviations from the mean, all participants were retained for the final analyses.

The data were analyzed using a mixed-design analysis of variance, with feedback type as the between-subjects factor and time as the within-subjects factor. Effect sizes were reported using partial eta squared. Assumptions of normality, homogeneity of variance, and sphericity were examined prior to analysis, and appropriate statistical corrections were applied where necessary.

Ethical considerations

All ethical guidelines for research involving human participants were observed throughout the study. No identifiable personal information was shared with the AI tool. Participants were fully informed about the

purpose of the study, the use of AI-generated feedback, and their right to withdraw at any stage without penalty.

Results

This section reports descriptive statistics and the results of the mixed-design ANOVA examining the effects of feedback type and time on learners' sentence-level accuracy in producing the English passive causative construction.

Descriptive statistics

Based on the scores from pre-test, post-test, and delayed post-test, the means and standard deviations for sentence-level accuracy have been presented in Table 1 for each feedback condition.

Table 1. Means and Standard Deviations of Accuracy Scores by Feedback Type and Time

Feedback type		Pretest	Posttest	Delayed posttest
AI Metalinguistic	Mean	1.4400	11.4800	10.2000
	N	25	25	25
	Std. Deviation	1.87261	3.57211	3.59398
AI Direct	Mean	1.6061	10.5758	7.3636
	N	33	33	33
	Std. Deviation	1.86982	3.58844	3.27699
Total	Mean	1.5345	10.9655	8.5862
	N	58	58	58
	Std. Deviation	1.85638	3.57852	3.67082

As shown in Table 1, both groups demonstrated very low and comparable accuracy on the pretest, indicating limited prior knowledge of the target structure. Marked improvements were observed for both groups on the immediate posttest. However, performance declined on the delayed posttest for both groups, with a larger decrease observed in the AI direct corrective feedback group.

Inferential statistics

Prior to inferential analysis, the data were screened for missing values and outliers. No missing data were identified, and standardized z-scores indicated no extreme outliers. Levene's tests confirmed that the assumption of homogeneity of variance was met at each testing occasion. However, Mauchly's test indicated that the assumption of sphericity was violated; therefore, Greenhouse–Geisser corrections were applied to the within-subjects effects. Given the robustness of repeated-measures ANOVA and the consistency of multivariate test results, the data were deemed suitable for further analysis.

To answer the research questions, a mixed-design analysis of variance (ANOVA) was conducted with feedback type (AI-generated metalinguistic feedback vs. AI-generated direct corrective feedback) as the between-subjects factor and time (pretest, posttest, delayed posttest) as the within-subjects factor.

As shown in Table 2, the analysis revealed a statistically significant main effect of time, $F(1.28, 72.09) = 540.37$, $p < .001$, $\eta^2 = .90$, indicating that learners' sentence-level accuracy changed significantly across the three testing occasions.

Table 2. Results of the Mixed-design ANOVA for Sentence-level Accuracy

Effect	F	p	η^2
Time	540.374	<.001	.906
Group (feedback type)	2.65	.109	.045
Time×Group	12.675	<.001	.185

The main effect of feedback type was not statistically significant, $F(1, 56) = 2.65$, $p = .109$, $\eta^2 = .045$, suggesting that when accuracy scores were averaged across the three testing times, the two feedback conditions did not differ significantly overall.

Importantly, the interaction between time and feedback type was statistically significant, $F(1.28, 72.09) = 12.67$, $p < .001$, $\eta^2 = .18$ (Table 2), indicating that the two feedback conditions exhibited different developmental paths over time.

Table 3. Estimated Marginal Means of Accuracy by Feedback Condition and Time

Feedback type	Time	Mean	Std. Error
AI	1	1.44	.37
Metalinguistic	2	11.48	.71
	3	10.20	.68
AI Direct	1	1.60	.32
	2	10.57	.62
	3	7.36	.59

As illustrated by the estimated marginal means in Table 3, both groups demonstrated marked improvement from pretest to posttest. However, learners who received AI-generated metalinguistic feedback showed less decline between the posttest and delayed posttest than those who received AI-generated direct corrective feedback. This pattern suggests a relative advantage for metalinguistic feedback in maintaining gains in accuracy over the three-week interval examined in this study.

Discussion

Given the expanding role of artificial intelligence in language teaching, particularly in the development of writing skills, this study examined the effects of two types of AI-generated written corrective feedback, direct corrective feedback and metalinguistic feedback, on EFL learners' grammatical accuracy over

time. Learners' sentence-level accuracy in producing the passive causative construction was assessed across three testing occasions. The results showed a significant main effect of time, indicating changes in grammatical accuracy across the three testing occasions. More importantly, a significant interaction between time and feedback type suggested that the two feedback conditions were associated with distinct developmental patterns rather than uniform learning outcomes.

The significant main effect of time indicates that learners' grammatical accuracy changed across the three testing occasions. Both feedback groups showed marked improvement from the pretest to the immediate posttest, followed by some decline on the delayed posttest, although performance remained above baseline levels. This pattern suggests that the gains observed immediately after treatment were only partially maintained over the three-week interval examined in the present study. The overall improvement following treatment is consistent with previous written corrective feedback research demonstrating that corrective feedback can facilitate grammatical development (Bitchener & Knoch, 2010). It also aligns with recent studies reporting improvements following AI-supported written corrective feedback (Abdi Tabari et al., 2025; Wang, 2024).

At the same time, the significant interaction between time and feedback type indicates that these gains did not develop in the same way across conditions. Although both groups showed marked improvement on the immediate posttest, learners who received AI-generated metalinguistic feedback demonstrated less decline in accuracy on the delayed posttest. This pattern suggests that metalinguistic feedback was associated with greater maintenance of grammatical accuracy over the three-week period for the target structure examined in this study. Similar variation in the relative effectiveness of feedback types has been reported in previous written corrective feedback research, with outcomes depending on the target structure and instructional conditions (Shintani & Ellis, 2013; Suzuki et al., 2019).

When accuracy scores were averaged across time, feedback type did not show an independent effect. This result highlights the importance of examining feedback effectiveness from a developmental rather than a cross-sectional perspective. Because the significant interaction showed that the two groups changed differently across the three testing occasions, the interaction effect provides a more informative account of feedback effectiveness than the main effect of feedback type alone.

The findings also extend research on AI-mediated written corrective feedback by demonstrating that the effectiveness of AI depends not only on the availability of automated feedback but also on how that feedback is designed. Rather than functioning as a single instructional approach, generative AI can deliver qualitatively different forms of corrective feedback that may be associated with different patterns of grammatical learning over time. These findings suggest that future research should move beyond evaluating AI as a single intervention and instead examine different AI feedback designs.

The present findings also complement emerging research on AI-generated written corrective feedback. Recent studies have shown that generative AI can provide adaptive and explanatory feedback during writing activities (EIEbyary & Shabara, 2024; Yang et al., 2025). The present study extends this work by

showing that different AI-generated feedback types may be associated with different developmental patterns in grammatical accuracy over time.

At the same time, the findings caution against assuming that one feedback type is universally superior across instructional contexts. Although both feedback conditions were associated with substantial immediate improvements in grammatical accuracy, direct corrective feedback appears particularly useful in contexts where immediate performance or rapid error correction is prioritized, such as drafting activities or time-constrained writing tasks. In contrast, the smaller decline observed in the metalinguistic feedback group suggests that explanatory feedback may be better suited to instructional goals emphasizing explicit understanding of the target structure and the maintenance of grammatical accuracy over time.

From a methodological perspective, the inclusion of both immediate and delayed posttests proved essential for interpreting feedback effectiveness. Studies relying solely on revision-based measures or immediate posttests risk conflating temporary performance gains with more stable learning outcomes. The divergent developmental patterns observed in the delayed posttest highlight the importance of including delayed measures when evaluating feedback effectiveness.

A further contribution of the present study lies in its focus on sentence-level production of a single grammatical structure. Unlike broader writing assessments, which combine grammatical accuracy with other dimensions of writing quality, this design allowed a more direct examination of changes in learners' control of the passive causative construction. Although this focus limits generalizability to overall writing performance, it strengthens the interpretation of differences between feedback conditions.

Overall, the findings suggest that AI-generated written corrective feedback can support grammatical learning, but its effectiveness depends on the characteristics of the feedback provided rather than the use of AI alone. The different developmental patterns observed across the two feedback conditions suggest that feedback explicitness remains an important consideration when integrating generative AI into L2 writing instruction. More broadly, the study suggests that insights from written corrective feedback research remain relevant for understanding the pedagogical potential of AI-assisted language learning.

Conclusion

The present study examined the effects of two types of AI-generated written corrective feedback on EFL learners' grammatical accuracy across pretest, immediate posttest, and delayed posttest measures. The findings showed that AI-mediated feedback, regardless of type, led to significant improvements in grammatical accuracy over time, highlighting the potential of generative AI as an effective medium for delivering form-focused instruction. At the same time, the results indicated that feedback type interacted with time, with learners in the two feedback conditions following different developmental paths, resulting in different levels of accuracy at the delayed posttest.

Taken together, the findings are consistent with previous research suggesting that the explicitness of written corrective feedback may influence the maintenance of grammatical accuracy over time. While direct corrective feedback resulted in strong immediate improvement, metalinguistic feedback was

associated with greater stability of accuracy over time. This pattern suggests that explanatory feedback may be particularly effective in supporting retention of grammatical knowledge over time. In this respect, the study both aligns with and extends earlier research on written corrective feedback and form-focused instruction by showing that these well-established principles remain applicable when feedback is delivered through AI rather than by human instructors.

The study, however, should be interpreted in light of several limitations. It was conducted with intact classes from a single institutional context, and participants' proficiency was inferred from their educational background rather than measured through a standardized proficiency test. In addition, the intervention focused on a single grammatical structure and examined retention over a three-week period following one treatment session. Future research should investigate the effects of AI-generated feedback across different learner populations, grammatical structures, and longer instructional periods, while also exploring how AI-generated and teacher-mediated feedback may be combined to support language development.

References

- Abdi Tabari, M., Kushki, A., & Wang, Y. (2025). Comparing the effects of teacher- and AI-mediated corrective feedback on accuracy, complexity, and quality in L2 written narratives. *Computer Assisted Language Learning*, 1–24. <https://doi.org/10.1080/09588221.2025.2561608>
- Ajabshir, Z. F., & Ebadi, S. (2023). The effects of automatic writing evaluation and teacher-focused feedback on CALF measures and overall quality of L2 writing across different genres. *Asian-Pacific Journal of Second and Foreign Language Education*, 8(1), 26. <https://doi.org/10.1186/s40862-023-00201-9>
- Bitchener, J., & Knoch, U. (2010). The contribution of written corrective feedback to language development: A ten-month investigation. *Applied Linguistics*, 31(2), 193–214. <https://doi.org/10.1093/applin/amp016>
- Cengiz, B. C., Bilki, Z., Ataş, A. H., & Celik, B. (2025). Exploring second language writers' engagement with ChatGPT feedback: Revision behaviors and perceptions. *System*, 134, 103837. <https://doi.org/10.1016/j.system.2025.103837>
- DeKeyser, R. M. (2007). Skill acquisition theory. In B. VanPatten & J. Williams (Eds.), *Theories in second language acquisition: An introduction* (pp. 97–113). Lawrence Erlbaum.
- EIEbyary, K., & Shabara, R. (2024). ChatGPT-generated corrective feedback: Does it do what it says on the tin? *Teaching English with Technology*, 24(3), 68–89. <https://doi.org/10.56297/vaca6841/BFFO7057/MYEH4562>
- Ellis, R. (2002). Does form-focused instruction affect the acquisition of implicit knowledge? A review of the research. *Studies in Second Language Acquisition*, 24(2), 223–236. <https://doi.org/10.1017/S0272263102002073>
- Ellis, R. (2009). A typology of written corrective feedback types. *ELT Journal*, 63(2), 97–107. <https://doi.org/10.1093/elt/ccn023>

- Ellis, R. (2015). Form-focused instruction and the measurement of implicit and explicit L2 knowledge. In P. Rebuschat (Ed.), *Implicit and explicit learning of languages* (pp. 417-441). John Benjamins.
- Ellis, N. (2015). Implicit AND explicit language learning: Their dynamic interface and complexity. In P. Rebuschat (Ed.), *Implicit and explicit learning of languages* (pp. 3-23). John Benjamins.
- Guo, K., & Wang, D. (2024). To resist it or to embrace it? Examining ChatGPT's potential to support teacher feedback in EFL writing. *Education and Information Technologies*, 29(7), 8435-8463. <https://doi.org/10.1007/s10639-023-12146-0>
- Kohnke, L., Moorhouse, B. L., & Zou, D. (2023). ChatGPT for language teaching and learning. *Relc Journal*, 54(2), 537-550. <https://doi.org/10.1177/00336882231162868>
- Li, S. (2010). The effectiveness of corrective feedback in SLA: A meta-analysis. *Language Learning*, 60(2), 309-365. <https://doi.org/10.1111/j.1467-9922.2010.00561.x>
- Lyster, R., & Ranta, L. (1997). Corrective feedback and learner uptake: Negotiation of form in communicative classrooms. *Studies in Second Language Acquisition*, 19(1), 37-66. <https://doi.org/10.1017/S0272263197001034>
- Mao, Z., Lee, I., & Li, S. (2024). Written corrective feedback in second language writing: A synthesis of naturalistic classroom studies. *Language Teaching*, 57(4), 449-477. <https://doi.org/10.1017/S0261444823000393>
- Mehrara, E., Ajideh, P., Zohrabi, M., Yaghubi Notash, M., & Hadidi, Y. (2025). Cultivating Opportunities: The Role of Guided Mobile-Assisted Task Performance on Reactivity to Mistake and L2 Willingness to Communicate. *Journal of Linguistic Studies: Theory and Practice*, 4(3), 43-66. <https://doi.org/10.22034/jls.2025.142706.1194>
- Muñoz, B. C. M., Nassaji, H., & Carrillo, F. I. B. (2025). ChatGPT-generated versus human direct corrective feedback on L2 writing. *System*, 134, 103805. <https://doi.org/10.1016/j.system.2025.103805>
- Rahmani, P., Zoghi, M., & Davatgari Asl, H. (2025). Ongoing application of peer and self-Assessment: A mixed-Methods study. *Journal of Linguistic Studies: Theory and Practice*. <https://doi.org/10.22034/jls.2025.143340.1231>
- Ranalli, J., Link, S., & 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. *Educational Psychology*, 37(1), 8-25. <https://doi.org/10.1080/01443410.2015.1136407>
- Rezazadeh, M., Tavakoli, M., & Eslami Rasekh, A. (2015). The effects of direct corrective feedback and metalinguistic explanation on EFL learners' implicit and explicit knowledge of English definite and indefinite articles. *Journal of English Language Teaching and Learning*, 7(16), 113-146. https://elt.tabrizu.ac.ir/article_17206.html
- Saricaoglu, A. (2019). The impact of automated feedback on L2 learners' written causal explanations. *ReCALL*, 31(2), 189-203. <https://doi.org/10.1017/S095834401800006X>

- Shintani, N., & Ellis, R. (2013). The comparative effect of direct written corrective feedback and metalinguistic explanation on learners' explicit and implicit knowledge of the English indefinite article. *Journal of Second Language Writing*, 22(3), 286–306. <https://doi.org/10.1016/j.jslw.2013.03.011>
- Shintani, N., Ellis, R., & Suzuki, W. (2014). Effects of written feedback and revision on learners' accuracy in using two English grammatical structures. *Language Learning*, 64(1), 103–131. <https://doi.org/10.1111/lang.12029>
- Shiyao, W. (2021). Understanding the effects of automated writing corrective feedback on L2 writing accuracy across proficiency levels. *Frontiers in Educational Research*, 4(11), 94–98. <https://doi.org/10.25236/FER.2021.041117>
- Suzuki, W., Nassaji, H., & Sato, K. (2019). The effects of feedback explicitness and type of target structure on accuracy in revision and new pieces of writing. *System*, 81, 135–145. <https://doi.org/10.1016/j.system.2018.12.017>
- Teng, F. M. (2024). A systematic review of ChatGPT for English as a foreign language writing: Opportunities, challenges, and recommendations. *International Journal of TESOL Studies*, 6(3), 36–57. <https://doi.org/10.58304/ijts.20240304>
- Truscott, J. (1996). The case against grammar correction in L2 writing classes. *Language Learning*, 46(2), 327–369. <https://doi.org/10.1111/j.1467-1770.1996.tb01238.x>
- Urzúa, C. A. C., Ranjan, R., Saavedra, E. E. M., Badilla-Quintana, M. G., Lepe-Martínez, N., & Philominraj, A. (2025). Effects of AI-assisted feedback via generative chat on academic writing in higher education students: A systematic review of the literature. *Education Sciences*, 15(10), 1396. <https://doi.org/10.3390/educsci15101396>
- Wang, D. (2024). Teacher-versus AI-generated (Poe application) corrective feedback and language learners' writing anxiety, complexity, fluency, and accuracy. *International Review of Research in Open and Distributed Learning*, 25(3), 37–56. <https://doi.org/10.19173/irrodl.v25i3.7646>
- Yan, D., & Zhang, S. (2024). L2 writer engagement with automated written corrective feedback provided by ChatGPT: A mixed-method multiple case study. *Humanities and Social Sciences Communications*, 11(1), 1–14. <https://doi.org/10.1057/s41599-024-03543-y>
- Yang, Y., Hussin, S., & Hashim, H. (2025). Revising with intelligence: ChatGPT feedback and its impact on EFL students' revision and self-efficacy. *Forum for Linguistic Studies*, 7(7), 352–367. <https://doi.org/10.30564/fls.v7i7.9845>
- Zhang, Y. (2024). Incorporating ChatGPT as an automated written corrective feedback tool into L2 writing class. *Journal of Language Teaching*, 4(4), 22–34. <https://doi.org/10.54475/jlt.2024.024>
- Zhang, Z. V., & Hyland, K. (2018). Student engagement with teacher and automated feedback on L2 writing. *Assessing Writing*, 36, 90–102. <https://doi.org/10.1016/j.asw.2018.02.004>