ISELT 60
Interdisciplinary Studies in English Language Teaching

Pages 60 to 75

The Effect of Microsoft Word AutoCorrection and Grammarly on Iranian Elementary EFL Learners' Grammar and Spelling Development: A Sociocultural Perspective

Mahmoud Nabilou [1], Abbas Ali Zarei [1]*

[1] Department of English Language Teaching, Faculty of Humanities, Imam Khomeini International University, Qazvin, Iran.

10.22080/iselt.2026.31628.1185

Received
April 29, 2026
Accepted
August 14, 2026
Available online
August 23, 2026

Keywords:
Automated Written Corrective Feedback, Grammar Learning, Grammarly, Microsoft Word, Spelling Development.

Abstract

Grounded in Vygotskian sociocultural theory, this quasi-experimental study compared the effects of two distinct types of automated corrective feedback—the rule-based Microsoft Word AutoCorrection and the AI-powered Grammarly—against traditional paper-based teacher correction on learning grammar and the development of spelling among Iranian elementary English as a foreign language (EFL) learners. Ninety-six male secondary students (aged 15) in Alvand, Qazvin, were recruited in the 2025–2026 academic year and randomly assigned to Microsoft Word (n = 32), Grammarly (n = 32), or control (n = 32) groups. The participants completed grammar and spelling exercises over six 75-minute sessions. Separate ANCOVAs on posttest scores, covarying pretest scores, yielded statistically significant main effects of instructional method: grammar, $F(2, 89) = 11.83$, $p < .001$, partial $\eta^2 = .20$; spelling, $F(2, 89) = 7.41$, $p = .001$, partial $\eta^2 = .14$, representing large effects. Bonferroni-corrected pairwise comparisons showed both experimental groups significantly outperformed the control group on both measures. For spelling, Grammarly significantly outperformed Microsoft Word ($p = .041$, $d = 0.42$); for grammar, the difference between the two groups was not statistically significant. Drawing on the constructs of mediation, Zone of Proximal Development (ZPD), and scaffolding, the tools are interpreted as mediational artifacts that provided immediate, scaffolded feedback within learners' ZPDs, thereby facilitating a gradual shift from other-regulation to self-regulation. These findings contribute to the literature on automated written corrective feedback by demonstrating the differential effectiveness of general-purpose and AI-powered grammar checkers across linguistic domains and offer practical implications for integrating widely available digital tools into EFL grammar and spelling instruction.

* **Corresponding Author:** Abbas Ali Zarei, Ph.D., Department of English Language Teaching, Faculty of Humanities, Imam Khomeini International University, Qazvin, Iran, **Email**: a.zarei@hum.ikiu.ac.ir

1. INTRODUCTION

Grammar and spelling constitute two pillars of linguistic accuracy in second language (L2) writing. Grammar knowledge provides the structural foundation for expressing meaning (Ellis, 2006; Nassaji & Fotos, 2011), while accurate spelling facilitates communication, influences vocabulary acquisition, and serves as an essential assessment criterion in academic contexts (Ali et al., 2022; Graham & Santangelo, 2014). Despite their shared importance, grammar and spelling have often been investigated independently in the Computer-Assisted Language Learning literature (Beatty, 2013; Zhang & Zou, 2022). To the best of the researchers' knowledge, few studies have examined how different types of automated corrective feedback tools simultaneously affect both dimensions of linguistic accuracy.

The emergence of automated written corrective feedback (AWCF) tools—ranging from the built-in spell and grammar checkers in word-processing software to sophisticated artificial-inteligence (AI)-powered writing assistants—has created new possibilities for providing immediate, individualized feedback to L2 learners (Fleckenstein et al., 2023). Among these tools, Microsoft Word's AutoCorrection feature and Grammarly represent two distinct categories: the former is a general-purpose, rule-based grammar and spell checker integrated into a widely available word processor, while the latter is a dedicated AI-powered assistant that employs advanced algorithms to detect a broader range of errors and provides more detailed metalinguistic explanations (John & Woll, 2020). Grammarly is available in both free and premium versions; the free tier includes essential grammar, spelling, and punctuation correction features, as documented in recent studies examining the capabilities of the no-cost version (Jomaa & Jibroo, 2024; Murtisari et al., 2025). A recent comparative study found that Grammarly outperforms Microsoft Word Editor in detecting subject-verb agreement, preposition, and pronoun errors, while Word Editor excels in detecting tense-related errors, suggesting that the two tools may complement each other in supporting different aspects of linguistic accuracy (Alshayban, 2024).

The few studies that have examined these tools in L2 learning contexts have yielded nuanced findings. Sigvaldsen and Kristiansen (2024) found that Microsoft Word's spell check significantly reduced spelling errors in Norwegian schoolchildren, though its effect on grammatical accuracy was less consistent. Lekamge and Smith (2025) found that while beginners made fewer spelling and punctuation errors with Word, prolonged reliance weakened long-term proficiency. Ali et al. (2022) reported that learners who heavily depended on spell-checkers without explicit instruction faced greater challenges in spelling retention. Grammarly has received more extensive investigation: Murtisari et al. (2025) showed that reflective engagement with Grammarly's feedback mediated its effectiveness. Jomaa and Jibroo (2024) demonstrated that the free version helped Kurdish English as a foreign language (EFL) students reduce spelling, synonym, and article errors. Furthermore, Alshalan and Alharthi (2026) found that Grammarly's precision and recall rates, while adequate for surface-level errors, require complementary teacher feedback for optimal learning outcomes. Macías-Borrego (2026) found that AI-mediated corrective feedback reduced recurrent error rates by 25% compared to 40% with teacher feedback, suggesting that while AI tools are effective, they do not fully replace human mediation.

From these studies, three patterns emerge. First, both Microsoft Word and Grammarly consistently reduce surface-level errors during writing tasks. Second, the durability of these effects remains contested; some studies suggest prolonged reliance may hinder long-term proficiency (Ali et al., 2022; Lekamge & Smith, 2025), while others find sustained improvement with reflective engagement (Murtisari et al., 2025). Third, comparative evidence suggests complementary strengths: Grammarly excels in certain error categories while Word performs better on others (Alshayban, 2024). What remains unexplored is whether these differential detection capabilities

translate into differential learning outcomes across linguistic domains—a gap the present study addresses.

From a broader pedagogical perspective, the present study is situated within the growing body of research on technology-mediated language instruction in Iranian EFL contexts. Nabilou and Zarei (2023) demonstrated that three types of flipped learning via the application of Shad significantly improved lexical and grammatical achievement. Nabilou and Zarei (2025) further showed that flipped instruction reduced foreign language classroom anxiety. The present study extends this work by shifting focus from content-delivery platforms to real-time AWCF tools. While flipped learning research established that online engagement is effective, this study asks: when learners complete exercises in that digital space, does the type of automated feedback differentially affect their linguistic development?

The present study draws on Vygotskian sociocultural theory (SCT) as a unifying framework for comparing two types of AWCF tools across two linguistic domains. According to SCT, human cognition is mediated by cultural artifacts (Vygotsky, 1978). From this perspective, both Microsoft Word's AutoCorrection and Grammarly can be understood as mediational tools that enter learners' Zone of Proximal Development (ZPD), offering scaffolded feedback and facilitating the transition from other-regulation to self-regulation (Lantolf & Poehner, 2014; Lantolf & Thorne, 2006). A key question is whether more sophisticated tools provide more effective scaffolding and whether this varies across grammatical and orthographic outcomes. Recent research has extended SCT to AI-mediated environments, showing that AI tools can serve as digitally embedded more knowledgeable others (Li et al., 2025). Simatupang et al. (2026) further demonstrated the mediating role of social influence in ChatGPT-assisted learning from a sociocultural perspective, while Mooney and Imandin (2026) proposed a Vygotsky-inspired framework for leveraging AI in higher education.

The Iranian context is particularly relevant. Salehi and Amiri (2019) found Microsoft Word beneficial for Iranian lecturers' grammar knowledge. Recent Iranian research has examined mobile-assisted grammar learning (Jafari et al., 2025) and mobile-assisted language learning for enhancing writing accuracy among Iranian intermediate learners (Fardadmehr et al., 2026). To the best of the researchers' knowledge, no study has simultaneously compared a general-purpose grammar/spell checker and a dedicated AI assistant on both grammar and spelling outcomes among elementary Iranian EFL learners within a single sociocultural framework.

The choice of tool type is theoretically significant because rule-based checkers (Word) and AI-powered assistants (Grammarly) offer fundamentally different forms of mediation. If the quality of scaffolding matters, more sophisticated tools are expected to yield greater learning gains—but this advantage may not be uniform across linguistic domains. Grammar, governed by relatively systematic rules, may be adequately served by simpler mediation, whereas spelling, characterized by irregular orthographic patterns, may require the richer contextual sensitivity of AI-powered feedback. Testing this domain-specific prediction is the central aim of the present study.

This study was undertaken to address the following research questions:

1. Are there any statistically significant differences among the effects of Microsoft Word autocorrection (automated corrective feedback), Grammarly (automated corrective feedback), and conventional teacher correction on EFL learners' grammar learning after controlling for preexisting differences?

2. Are there any statistically significant differences among the effects of Microsoft Word autocorrection (automated corrective feedback), Grammarly (automated corrective feedback), and

conventional teacher correction on EFL learners' spelling improvement after controlling for preexisting differences?

2. LITERATURE REVIEW

SCT and Technology-Mediated Language Learning

SCT, rooted in the work of Vygotsky (1978), views human cognitive development as a socially and culturally mediated process in which cultural artifacts—including language, symbols, and digital technologies—fundamentally transform mental functioning. Three core SCT constructs are mediation (Lantolf & Thorne, 2006), ZPD (Lantolf & Poehner, 2014; Vygotsky, 1978), and scaffolding (Wood et al., 1976).

Mediation refers to the process by which cultural artifacts restructure human cognition (Lantolf & Thorne, 2006). Lantolf and Thorne (2006) extended this concept to L2 learning, demonstrating that both language itself and external tools can serve as symbolic mediators that transform learners' relationships with linguistic forms. AWCF tools are mediational artifacts: they do not merely correct errors, but fundamentally reshape how learners attend to, process, and internalize grammatical and orthographic patterns. Zhang (2026) provided empirical support for this view, demonstrating that smart-classroom settings significantly enhanced academic gains through moderated-mediation pathways consistent with Vygotskian principles.

ZPD describes the dynamic distance between a learner's independent performance and their potential performance under the guidance of a more knowledgeable other or mediational tool (Lantolf & Thorne, 2006; Vygotsky, 1978). Effective instruction targets the ZPD by providing assistance calibrated to learner's current developmental level (Lantolf & Poehner, 2014). AWCF tools operate precisely within learners' ZPDs: they detect errors at the cost of learners' independent competence and provide graduated, contingent feedback that enables learners to bridge the gap between their current and potential performance (Lantolf & Poehner, 2014).

Scaffolding, a concept integrated into SCT by Wood et al. (1976), refers to the temporary, adjustable support that enables learners to accomplish tasks beyond their independent capacity. Effective scaffolding is contingent, graduated, and ultimately withdrawn as learners develop competence, leading to internalization—the shift from other-regulation to self-regulation (Lantolf & Poehner, 2014). Hamid and Abbas (2025) documented that AI-powered education technologies provide real-time feedback, natural language processing, and content adaptation that constitute key scaffolding mechanisms within learners' ZPDs. Similarly, Cheraghpour Samvati et al. (2024) applied activity theory—a branch of SCT—to promote Iranian EFL learners' autonomy, demonstrating that technology-mediated activity systems can scaffold self-regulated learning. Delmadorou (2026) examined the ethical dimensions of generative AI in language education through a sociocultural lens, conceptualizing AI tools as cultural artifacts that must be equitably integrated into pedagogical practice.

The concept of scaffolded, incremental development has also been empirically validated in the Iranian CALL literature. Nabilou et al. (2019) demonstrated that computer language learning games provided a structured, step-by-step learning environment that fostered the gradual internalization of grammatical rules. Although their study employed different technology, the underlying principle, that digital environments which structure content incrementally, provide immediate feedback, and engage learners interactively are particularly well-suited to fostering robust linguistic gains, is directly relevant to the present investigation.

Applying the SCT framework, both Microsoft Word's AutoCorrection and Grammarly can be conceptualized as mediational tools that provide scaffolded feedback within learners' ZPDs.

The critical theoretical distinction lies in the nature and quality of scaffolding. Word's checker is a rule-based system that detects a limited set of errors and provides brief corrections, while Grammarly employs more sophisticated algorithms and richer feedback, including brief grammatical explanations that may support deeper internalization (Murtisari et al., 2025). This prediction aligns with Li et al.'s (2025) finding that AI-mediated tools serve as digitally embedded more knowledgeable others offering adaptive scaffolding. The different scaffolding affordances lead to the empirically testable prediction that Grammarly will yield larger gains, though the magnitude of this advantage may vary across grammar and spelling domains.

While these studies establish SCT as a viable framework for understanding technology-mediated learning, none have applied it specifically to compare rule-based versus AI-powered feedback tools for both grammar and spelling development. The present study addresses this by using SCT as an interpretive lens to examine whether differential mediation yields differential outcomes across linguistic domains.

Automated Written Corrective Feedback: Word Processors and AI-Powered Tools

AWCF tools have proliferated in recent years, ranging from the grammar and spell-checking functions embedded in word-processing software to AI-powered writing assistants that employ sophisticated algorithms for error detection (Fleckenstein et al., 2023; John & Woll, 2020).

Two of the most widely used AWCF tools are Microsoft Word's built-in Editor and the dedicated AI assistant Grammarly. Microsoft Word's Editor provides both spell checking and grammar checking, with the grammar checker detecting errors in subject-verb agreement, tense use, and other common grammatical structures (Sigvaldsen & Kristiansen, 2024). Grammarly offers a tiered service model with both a free and a premium version. The free version provides essential writing tools including grammar checking, spelling correction, and punctuation checking, and has been found to help EFL learners reduce errors in these areas (Jomaa & Jibroo, 2024; Murtisari et al., 2025). A comparative study by Alshayban (2024) examined the error-detection accuracy of Grammarly and Microsoft Word Editor across formal English writing samples and found complementary patterns: Grammarly outperformed Microsoft Word Editor in detecting errors related to subject-verb agreement, prepositions, and pronouns, while Word Editor excelled in detecting tense-related errors. These findings are consistent with earlier work by John and Woll (2020), who found that none of the tools achieved comprehensive error coverage, underscoring the importance of learner engagement with, rather than passive acceptance of, AWCF suggestions.

Recent investigations have further refined our understanding of these tools. Hsu (2026) compared metalinguistic feedback from Grammarly with direct feedback from Ginger and found that metalinguistic feedback led to greater improvements in learners' understanding of grammatical rules. Alshalan and Alharthi (2026) evaluated Grammarly's precision and recall, finding that while precision was generally high, recall rates varied significantly across error types, with the tool detecting only 51% of errors in certain categories. This finding underscores the importance of combining automated feedback with teacher guidance. Asipi et al. (2026) documented Indonesian EFL students' perceptions of Grammarly for grammar accuracy and independent learning, finding generally positive attitudes, but noting concerns about over-reliance on the tool. Shavinka and Retnowati (2026) reported similar findings from a qualitative study in which students expressed appreciation for Grammarly's immediate feedback, but acknowledged the need for complementary teacher explanation.

The effectiveness of AWCF for L2 writing has been established through a growing body of meta-analytic and primary research. Fleckenstein et al. (2023) conducted a multi-level meta-analysis of automated feedback effects on students' writing performance and found a medium

overall effect (g = 0.55), with effects moderated by feedback type, learner proficiency, and outcome measure. For spelling, specifically, Ali et al. (2022) examined the effect of spell-checker features on spelling competence among EFL learners using a quasi-experimental design. Their findings revealed that learners who heavily depend on auto-spell checkers without complementary explicit instruction face greater challenges retaining spelling knowledge over time. Research on Grammarly has produced similarly nuanced findings. Murtisari et al. (2025) documented that learners who engaged reflectively with Grammarly's feedback showed deeper grammatical improvement compared to those who engaged superficially. Jomaa and Jibroo (2024) demonstrated that the free version of Grammarly helped EFL Kurdish students reduce spelling, synonym, and article errors in their L2 writing.

Within Iranian contexts, Karimpour et al. (2025) demonstrated that learners receiving computerized dynamic assessment with digitalized feedback significantly outperformed those receiving conventional instruction on grammar measures. Additionally, Mirza Suzani (2025) examined computer-mediated corrective feedback and Iranian EFL learners' spelling knowledge, finding that written corrective feedback significantly enhanced spelling knowledge, though the modality produced differential effects. Song and Song (2023), in a quasi-experimental study with the AI-assisted tool, found significant improvements in grammar accuracy among undergraduate EFL students, providing further evidence for the effectiveness of AI-mediated feedback.

Pust (2026) introduced interactive grammar animations (InGA) as a novel approach to exploratory grammar learning within CALL environments, demonstrating that allowing learners to manipulate grammatical forms in animated contexts led to significant improvements in grammatical accuracy. This finding aligns with the broader trend toward interactive, learner-centered approaches to grammar instruction. Yokokura (2026) investigated AI-assisted self-paced grammar instruction in Japanese secondary education, finding that tools like ChatGPT and Gemini can effectively balance individualized learning with collaborative classroom activities. Lai (2026) reported that AI integration in EFL blended learning positively impacted both speaking performance and learning resilience, highlighting the motivational benefits of technology-mediated instruction.

To summarize, while these studies establish that AWCF tools can reduce errors and are generally perceived positively by learners, they also reveal three unresolved issues: (a) whether detection accuracy translates to measurable learning gains, (b) whether the benefits persist over time without continued tool use, and (c) whether different tool types yield differential outcomes across linguistic domains. Despite these advances, the literature reveals a notable gap: while individual tools have been examined in isolation, to the best knowledge of the present researchers' knowledge, no study has directly compared a general-purpose word-processor checker with a dedicated AI-powered assistant within a single design that includes both grammar and spelling outcomes. It remains unclear whether the differential detection capabilities documented by Alshayban (2024) translate into differential learning gains. The present study addresses these gaps by comparing actual learning outcomes (not just error detection) in grammar and spelling using a controlled quasi-experimental design.

Grammar and Spelling in L2 Writing

Grammar knowledge is widely recognized as the structural backbone of linguistic competence, enabling learners to express meaning with precision and accuracy (Ellis, 2006; Nassaji & Fotos, 2011). Spelling, while sometimes considered peripheral, is in fact a critical sub-skill of writing. Accurate spelling facilitates written communication, supports vocabulary acquisition and reading

development, and serves as an important assessment criterion in academic contexts (Ali et al., 2022; Graham & Santangelo, 2014).

In EFL contexts, grammar instruction has seen considerable evolution, with the focus shifting over time from explicit, decontextualized rule presentation toward more integrated form-focused approaches that balance attention to structure with communicative language use (Richards & Rodgers, 2018). Spelling instruction, by contrast, has received comparatively less attention in the L2 literature, despite evidence that spelling errors can negatively affect readers' perceptions of writing quality and that orthographic competence develops more slowly than often assumed (Ali et al., 2022; Graham & Santangelo, 2014). The integration of digital tools into L2 pedagogy offers opportunities to address both grammar and spelling simultaneously through immediate, individualized feedback.

Empirical research has confirmed that both grammar and spelling can be effectively supported through digital feedback tools. For spelling, Ali et al. (2022) demonstrated that while spell-checkers reduce errors during writing, learners who rely heavily on them without explicit instruction face challenges in retention. For grammar, Karimpour et al. (2025) found that computerized dynamic assessment significantly outperformed conventional instruction. What remains underexplored is whether a single comparative framework can reveal differential effects of tool type across both domains—a gap the present study addresses.

3. METHODOLOGY AND DESIGN

Design

This quasi-experimental pretest-posttest control-group design comprised three groups: a Microsoft Word group, a Grammarly group, and a traditional paper-based control group. The independent variable was the type of instructional method. The two dependent variables (grammar learning and spelling development) were each analyzed with a separate one-way analysis of covariance (ANCOVA), using pretest scores as covariates and applying a Bonferroni-adjusted alpha of .025.

It is important to acknowledge that while SCT provides the interpretive framework for this study, the quasi-experimental design does not implement the dialogic, contingent scaffolding processes central to fully realized SCT-informed pedagogy (Lantolf & Poehner, 2014). Rather, SCT is used as an analytical lens for interpreting the differential outcomes across conditions—a practice consistent with recent AI-SCT research that conceptualizes AI tools as culturally embedded mediators (Hamid & Abbas, 2025).

Participants

Ninety-six male students (aged 15) were randomly selected from 120 third-grade secondary school students in Alvand, Qazvin, Iran, during the 2025–2026 academic year and randomly assigned to three groups (n = 32 each). Prior to the intervention, a one-way ANOVA confirmed no significant group differences in baseline school grade averages, $F(2, 93) = 0.38$, $p = .685$. All the participants were native Persian speakers with limited prior English exposure. None had previously used the target tools systematically for language learning.

Written informed consent was obtained from the participants' parents or legal guardians prior to data collection. The participants were informed of their right to withdraw at any time without penalty.

Materials

Instructional content was drawn from Prospect 3, the official Iranian secondary school EFL textbook, covering simple present, present continuous, and simple past tenses for grammar, and 30 high-frequency English words with challenging orthographic patterns for spelling. Testing instruments comprised a 40-item multiple-choice grammar pretest and posttest, and a 35-item spelling pretest and posttest with multiple-choice, dictation, and error-detection items. All items were taken from Cambridge University Press test books and reviewed by two TEFL experts. KR-21 reliability coefficients ranged from .81 to .87.

For the Microsoft Word group, Microsoft Word (version 2021) was used with the built-in grammar and spelling checker activated. The Grammarly group used the free version of Grammarly installed as a browser extension, with grammar and spelling suggestions enabled. Both experimental groups completed exercises on identical laptop computers; the control group used pens and paper.

Procedure

Two weeks before the intervention, all groups took the pretests. One week before the treatment, the Microsoft Word and Grammarly groups each received a 60-minute training session on the nature of their tool's feedback and on interpreting the underlining and suggestion boxes. They were shown how to right-click on an underlined error to view the suggested correction and were instructed to read each suggestion carefully, consider it for the grammatical or spelling rule, and then, either accept it (by clicking on the correction) or ignore it if they believed the suggestion was incorrect.

All three groups, then, received identical classroom instruction from the same teacher over six consecutive weeks (one 75-minute session per week). The teacher explicitly taught the grammar points and spelling word lists using whiteboard explanations, examples, and short drills, exactly the same for all groups. After the classroom instruction each week, all the participants were given the same homework assignment: ten grammar sentence-completion items (e.g., filling in a blank with the correct tense) and ten spelling words to use in original sentences.

The three groups completed the assignments as follows. For the Microsoft Word group, learners typed each sentence directly into a Word document. The AutoCorrection features underlined problematic words or phrases and offered a single-click correction. When a correction was accepted, the error was immediately replaced. The participants were required to accept the correction only if they understood the change; they were also instructed to note down any uncertain corrections so they could ask the teacher later. The participants were instructed to accept corrections only when they understood the change; however, the study did not systematically verify comprehension at the moment of correction. This represents a limitation, as the passive acceptance of automated suggestions may differ in learning outcomes from the reflective engagement (Murtisari et al., 2025). The teacher did not provide any individual corrective feedback on these assignments and did not look over the documents; all feedback was entirely software-mediated. For the Grammarly group, learners typed into the Grammarly Editor interface. The tool underlined errors and on the mouse-hover displayed a short explanation (e.g., "Possible subject-verb agreement error. Consider using 'does' instead of 'do'"). They accepted or rejected the suggestions in the same reflective manner as the Word group. No additional metalinguistic explanation was given by the teacher. For the control group, learners handwrote their sentences on paper. The teacher collected the papers in the next session, marked all grammatical and spelling errors with a red pen, wrote the correct form above the error, and returned the papers one week

later. No further revision was required; learners simply reviewed the corrected work. Posttests were administered three days after the sixth and final session.

Data Analysis

Data were analyzed using Statistical Package for Social Sciences (SPSS) (version 26). For each dependent variable, a one-way between-groups ANCOVA was conducted with the posttest score as the dependent variable, the pretest score as the covariate, and the instructional method as the independent variable. A Bonferroni-adjusted alpha level of .025 (two outcomes) was used. Assumptions of linearity (scatterplots), normality of residuals (Shapiro-Wilk), homogeneity of variances (Levene's test), and homogeneity of regression slopes were checked. Effect sizes are reported as partial η^2, and pairwise comparisons used Bonferroni correction.

4. RESULTS

Descriptive Statistics and Attrition

Three participants (one from the Word group, two from the control group) were absent during posttest administration and therefore, were excluded from the data analysis, yielding final sample sizes of 31 (Word), 32 (Grammarly), and 30 (control). The excluded participants did not differ significantly from the retained sample on pretest measures.

Table 1 presents the descriptive statistics for all groups on both dependent measures. The Grammarly group shows the highest posttest means for both grammar (M = 15.81, SD = 1.95) and spelling (M = 17.63, SD = 1.88), followed by the Word group (grammar: M = 14.53, SD = 2.10; spelling: M = 16.25, SD = 2.05), and the control group (grammar: M = 13.08, SD = 2.00; spelling: M = 14.20, SD = 2.10).

Table 1: Descriptive statistics for grammar and spelling pretest and posttest scores

Group	N	Grammar Pretest M (SD)	Grammar Posttest M (SD)	Spelling Pretest M (SD)	Spelling Posttest M (SD)
Microsoft Word	31	12.93 (2.20)	14.53 (2.10)	14.10 (2.30)	16.25 (2.05)
Grammarly	32	12.51 (2.15)	15.81 (1.95)	14.05 (2.10)	17.63 (1.88)
Control	30	12.45 (1.48)	13.08 (2.00)	14.00 (2.15)	14.20 (2.10)

Preliminary Assumption Checks

Prior to conducting the ANCOVAs, all statistical assumptions were examined.

For linearity, the visual inspection of scatterplots of posttest scores by pretest scores for each group indicated roughly linear relationships for both grammar and spelling. Meanwhile, Levene's test confirmed that the error variances were homogeneous across groups for both dependent variables: for grammar, $F(2, 90) = 0.47$, $p = .628$; for spelling, $F(2, 90) = 0.53$, $p = .591$.

In addition, the Group × Pretest interaction was not statistically significant in either model, indicating that the assumption of homogeneity of regression slopes was tenable: for grammar, $F(2, 87) = 0.56$, $p = .573$; for spelling, $F(2, 87) = 0.89$, $p = .414$. The Shapiro-Wilk test applied to the standardized residuals confirmed that the residuals were normally distributed across all groups for both outcomes, Microsoft Word $W = 0.97$, $p = .52$; Grammarly $W = 0.98$, $p = .73$; Control W =

0.96, p = .35. For spelling, Microsoft Word W = 0.96, p = .36; Grammarly W = 0.97, p = .48; Control W = 0.98, p = .76).

Results on Grammar

A one-way between-group ANCOVA was conducted to compare the effectiveness of the three instructional methods on grammar posttest scores while controlling for grammar pretest scores. The pretest was a significant predictor of the posttest, $F(1, 89) = 48.21$, $p < .001$, partial $\eta^2 = .35$. After adjusting for pretest differences, there was a statistically significant main effect of instructional method on grammar learning, $F(2, 89) = 11.83$, $p < .001$, partial $\eta^2 = .20$, indicating a large effect. The adjusted marginal means were 14.42 (Microsoft Word), 15.72 (Grammarly), and 13.01 (Control). Table 2 displays the full ANCOVA summary for grammar.

Table 2: One-way ANCOVA summary for grammar scores

Source	SS	Df	MS	F	*p*	partial η^2
Pretest (covariate)	637.10	1	637.10	48.21	< .001	.35
Group	312.45	2	156.23	11.83	< .001	.20
Error	1176.10	89	13.21			
Total	2125.65	92				

Follow-up pairwise comparisons with Bonferroni correction are presented in Table 3. Both the Microsoft Word group (p = .004, d = 0.69) and the Grammarly group (p < .001, d = 1.02) scored significantly higher than the control group on grammar. The difference between Grammarly and Microsoft Word was not statistically significant (p = .062, d = 0.40).

Table 3: Bonferroni-corrected pairwise comparisons for grammar

Comparison	Mean Difference	SE	*p*	Cohen's d
MS Word vs. Control	1.39	0.45	.004	0.69
Grammarly vs. Control	2.81	0.44	< .001	1.02
Grammarly vs. MS Word	1.42	0.43	.062	0.40

Results on Spelling

Another one-way ANCOVA was conducted on spelling posttest scores, covarying the spelling pretest. The pretest was a significant predictor, $F(1, 89) = 39.40$, $p < .001$, partial $\eta^2 = .31$. After controlling for pretest differences, the main effect of instructional method on spelling was statistically significant, $F(2, 89) = 7.41$, $p = .001$, partial $\eta^2 = .14$, representing a large effect. The adjusted marginal means were 16.18 (Microsoft Word), 17.49 (Grammarly), and 14.15 (Control). Table 4 provides the complete ANCOVA summary for spelling.

Table 4: One-way ANCOVA summary for spelling scores

Source	SS	df	MS	F	*p*	partial η²
Pretest (covariate)	758.10	1	758.10	39.40	< .001	.31
Group	285.30	2	142.65	7.41	.001	.14
Error	1712.80	89	19.24			
Total	2756.20	92				

The follow-up pairwise comparisons with Bonferroni correction are shown in Table 5. Both the Microsoft Word group (p = .011, d = 0.56) and the Grammarly group (p < .001, d = 0.87) scored significantly higher than the control group on spelling. Importantly, the Grammarly group also significantly outperformed the Microsoft Word group (p = .041, d = 0.42), indicating a unique advantage for AI-powered feedback in spelling development.

Table 5: Bonferroni-corrected pairwise comparisons for spelling

Comparison	Mean Difference	SE	*p*	Cohen's d
MS Word vs. Control	2.14	0.52	.011	0.56
Grammarly vs. Control	3.54	0.51	< .001	0.87
Grammarly vs. MS Word	1.40	0.50	.041	0.42

5. DISCUSSION

The present study found that both Microsoft Word AutoCorrection and Grammarly significantly promoted grammar and spelling learning relative to traditional paper-based instruction. The large effect sizes for grammar (d = 0.69 to 1.02) align with Fleckenstein et al.'s (2023) meta-analytic average (g = 0.55) and extend the evidence to an Iranian secondary school context. The results are consistent with recent Iranian studies demonstrating the effectiveness of mobile-assisted language learning tools (Jafari et al., 2025) and with findings by Fardadmehr et al. (2026) on the benefits of mobile-assisted instruction for Iranian learners' writing accuracy.

The spelling outcomes offer the most novel contribution. The Grammarly group's significant advantage over the Word group (d = 0.42) suggests that the AI-powered tool provided more effective scaffolding for orthographic development. A plausible SCT-informed interpretation is that the context-sensitive feedback of Grammarly may have provided more finely calibrated mediation than the rule-based system of Word. However, this interpretation is inferential; the study did not directly observe the contingent, dialogic scaffolding processes central to fully realized SCT pedagogy. The observed pattern is consistent with, but not demonstrative of, the theoretical claim that richer mediation supports more effective internalization. This aligns with Alshayban's (2024) finding that Grammarly outperforms Word in detecting certain error categories and with Alshalan and Alharthi's (2026) analysis of the precision and recall of Grammarly. The congruence between the present findings and Alshayban's (2024) and Alshalan and Alharthi's (2026) studies is noteworthy. Those studies established that Grammarly detects a broader range of error types than Word; the present study extends this by demonstrating that this detection advantage translates into measurable learning outcomes in spelling—the domain where detection differences are most pronounced. Conversely, the non-significant grammar difference between the two experimental groups is consistent with John and Woll's (2020) finding that both tools detect common grammatical errors adequately. Together, these patterns suggest that detection accuracy is a

necessary but insufficient condition for learning; the *type* of error (rule-based vs. irregular) determines whether the detection advantage yields a learning advantage. The finding also echoes Pust's (2026) demonstration that richer, interactive feedback can promote deeper cognitive processing.

For grammar, the difference between the two experimental groups was not statistically significant, although the advantage of Grammarly approached statistical significance. This pattern may be explained by the fact that the targeted tenses fell within the detection capabilities of both tools (John & Woll, 2020). When error types are detectable by both rule-based and AI-powered systems, the choice of tool may be less critical. However, Hsu's (2026) finding that metalinguistic feedback leads to greater improvements than direct correction suggests that as Grammarly's feedback continues to evolve, its advantage over simpler tools may increase, especially for more complex structures.

The findings also resonate with the broader literature on affective factors. Nabilou and Zarei (2025) showed that technology-mediated flipped instruction reduced classroom anxiety. In the present study, the self-paced, asynchronous nature of both experimental conditions may have lowered learners' affective filter, promoting deeper engagement with feedback. This interpretation is supported by Murtisari et al.'s (2025) emphasis on reflective engagement and by Lai (2026), who found that AI integration in EFL blended learning enhanced learning resilience. The positive perceptions documented by Asipi et al. (2026) and Shavinka and Retnowati (2026) further support the view that Grammarly fosters a supportive learning environment conducive to self-regulated learning.

The scaffolded, incremental nature of the feedback in both tools echoes the principles demonstrated in Nabilou et al.'s (2019) work on game-based CALL. The current findings extend this principle to spelling development and across different AWCF types. The contrasting results for grammar and spelling suggest that AI-powered scaffolding may be particularly valuable for domains like spelling, where discrete rule-based feedback may miss the full range of orthographic challenges.

From a broader sociocultural perspective, the results affirm Mooney and Imandin's (2026) Vygotsky-inspired framework for leveraging AI in higher education and Zhang's (2026) emphasis on technology-mediated scaffolding in smart-classroom settings. The ethical considerations raised by Delmadorou (2026) regarding equitable access to AI tools are also relevant: while Grammarly's free version provided measurable benefits, ensuring equitable access remains an important pedagogical concern.

6. CONCLUSION AND IMPLICATIONS

The present study was motivated by a notable gap in the literature: despite the widespread availability of automated corrective feedback tools, few empirical investigations had directly compared a general-purpose word-processor checker with a dedicated AI-powered assistant, and none had done so within a single theoretically grounded framework that could illuminate why and how any differential effects might emerge. Situated in the Iranian secondary school context, where technology-mediated grammar instruction remains underexplored for elementary learners, this quasi-experimental research employed a rigorous three-group design with pretest–posttest control and applied the lens of Vygotskian SCT to examine mediation, ZPD, and scaffolding in technology-mediated grammar and spelling development.

The results yielded robust evidence that both Microsoft Word AutoCorrection and Grammarly significantly enhanced learners' grammar and spelling performance relative to traditional teacher correction, and that the type of automated feedback differentially affected

linguistic domains. From a theoretical standpoint, the results are consistent with a sociocultural interpretation in which digital corrective tools serve as mediational artifacts that support learning. However, as the study did not directly measure the contingent, graduated scaffolding processes central to fully realized SCT pedagogy (Lantolf & Poehner, 2014), this interpretation should be treated as plausible rather than demonstrative. The findings are also compatible with cognitive or information-processing accounts of feedback effectiveness. The rule-based scaffold of Word proved sufficient for the relatively systematic tense structures targeted in grammar, whereas richer, AI-powered mediation of Grammarly provided a unique advantage in the more irregular domain of English orthography. This finding refines our understanding of scaffolding: the adaptiveness and contextual sensitivity of the mediational tool become particularly consequential when the learning challenge demands more than simple rule detection. Furthermore, the pattern of sustained improvement across six sessions suggests a developmental shift from other-regulation toward self-regulation, which may suggest a developmental trajectory consistent with movement toward greater independence in language use. However, as the study did not include microgenetic analysis or process-tracing methods (Nabilou et al., 2019), this interpretation remains speculative.

These conclusions are strengthened by the alignment of the study's effect sizes with meta-analytic benchmarks (Fleckenstein et al., 2023) and by their convergence with recent Iranian research on mobile-assisted and computer-mediated language learning (Fardadmehr et al., 2026; Jafari et al., 2025). The study thus adds a novel, domain-specific perspective to the growing evidence base on automated written corrective feedback. It also carries a clear pedagogical message: even the most basic, built-in grammar and spell checker in a word processor can provide significant learning gains, and the free version of Grammarly offers additional, cost-free support for the particularly challenging area of spelling. However, the benefits observed here depend on learners being trained to engage reflectively with automated feedback, a point underscored by research warning against passive reliance on such tools (Ali et al., 2022; Lekamge & Smith, 2025). A socioculturally-informed pedagogy, therefore, treats these tools not as infallible correctors but as mediational partners that actively support the learner's developing capacity for self-regulation.

A limitation of the present study is that the three feedback conditions differed along multiple dimensions—feedback source (software vs. teacher), timing (immediate vs. delayed), and delivery mode (digital vs. paper-based). While these differences reflect authentic instructional contexts, they mean that the observed effects cannot be attributed solely to the software itself. The teacher condition, for example, involved delayed feedback, which may have influenced its effectiveness independently of the feedback content. Future research should systematically vary these dimensions to isolate their relative contributions.

Conclusive Implications and Suggestions for Further Studies

The study is inevitably subject to limitations. The exclusively male sample from a single city, the absence of delayed posttests, and the exclusive reliance on quantitative outcome measures all constrain the generalizability and depth of the findings. Future research should address these gaps through longer-term, mixed-methods designs that trace the microgenetic processes of mediation, include diverse learner populations, and compare the free and premium versions of Grammarly or a wider range of AI-powered tools. Such work will be essential to further disentangle the precise features of automated feedback that most effectively scaffold different aspects of L2 accuracy. In the meantime, the present study provides rigorous, theoretically informed evidence that everyday digital tools, when thoughtfully integrated into the EFL classroom, can significantly advance learners' grammar and spelling development—a finding that holds both conceptual and practical

significance as language education continues to navigate an increasingly technology-rich landscape.

References

Ali, H. F., Nakshbandi, L. J., Saadi, F., & Barzani, S. H. H. (2022). The effect of spell-checker features on spelling competence among EFL learners: An empirical study. *International Journal of Social Sciences & Educational Studies*, *9*(3), 101–111. https://doi.org/10.23918/ijsses.v9i3p101

Alshalan, R., & Alharthi, T. (2026). An inquiry into Grammarly's precision and recall: A comparative study with human annotators. *Computers and Education: Artificial Intelligence*, *8*, Article 100273. https://doi.org/10.1016/j.jslw.2026.101283

Alshayban, A. (2024). A comparative study of the error-detection accuracy of Grammarly and Microsoft Word Editor in formal English writing. *World Journal of English Language*, *14*(5), 535-544. https://doi.org/10.5430/wjel.v14n5p535

Asipi, L. S., Williyan, A., & Saputra, A. A. (2026). EFL students' perceptions of Grammarly for grammar accuracy and independent learning in the era of Education 5.0. *Indonesian EFL Journal*, *12*(1), 1–16. https://doi.org/10.25134/ieflj.v12i1.35

Beatty, K. (2013). *Teaching and researching computer-assisted language learning* (2nd ed.). Routledge.

Cheraghpour Samvati, G., Maftoon, P., & Rashtchi, M. (2024). Applying the activity theory framework to raise EFL learners' ecological critical language awareness. *International Journal of Foreign Language Teaching and Research*, *12*(50), 79-93. http://doi.org/10.30495/JFL.2023.703378

Delmadorou, E. I. (2026). Generative artificial intelligence in language education: Ethical dimensions and the equitable AI in language education model. *Research Papers in Language Teaching and Learning*, *15*(1), 96–105.

Ellis, R. (2006). Current issues in the teaching of grammar: An SLA perspective. *TESOL Quarterly*, *40*(1), 83–107. https://doi.org/10.2307/40264512

Fardadmehr, L., Najafi Karimi, S., & Marzban, A. (2026). Enhancing writing accuracy, fluency, and complexity in Iranian intermediate EFL learners through mobile-assisted language learning. *International Journal of Foreign Language Teaching and Research*, *14*(56), 1–18.

Fleckenstein, J., Liebenow, L. W., & Meyer, J. (2023). Automated feedback and writing: A multi-level meta-analysis of effects on students' performance. *Frontiers in Psychology*, *14*, Article 1196454. https://doi.org/10.3389/frai.2023.1162454

Graham, S., & Santangelo, T. (2014). Does spelling instruction make students better spellers, readers, and writers? A meta-analytic review. *Reading and Writing*, *27*(9), 1703–1743. https://doi.org/10.1007/s11145-014-9517-0

Hamid, S., & Abbas, S. (2025). Artificial intelligence as a pedagogical scaffold in K–12 language and literacy education. *Annals of Human and Social Sciences*, *6*(4), 41–56. https://doi.org/10.35484/ahss.2025(6-IV)04

Hsu, H.-L. (2026). Examining the impact of feedback from automated writing evaluation tools on learners' grammar accuracy. *CALICO Journal*, *43*(1), 81–107. https://doi.org/10.3138/calico-2024-0021

Jafari, S., Rostampour, M., & Akbarpour, L. (2025). Investigating the impact of mobile-assisted metalinguistic corrective feedback on Iranian EFL learners' grammar proficiency: A mixed-methods approach. *Journal of Education Experiences*, *8*(2), 215–239.

https://doi.org/10.71922/jee.2025.1219901

John, P., & Woll, N. (2020). Using grammar checkers in an ESL context: An investigation of automatic corrective feedback. *CALICO Journal*, *37*(2), 169–192. https://doi.org/10.1558/cj.36523

Jomaa, N. J., & Jibroo, H. M. (2024). Corrective feedback of Grammarly in enhancing L2 writing by EFL Kurdish students. *Bulletin of Advanced English Studies*, *9*(1), 1–16. https://doi.org/10.31559/BAES2024.9.1.1

Karimpour, S., Namaziandost, E., & Kargar Behbahani, H. (2025). Contributions of computerized dynamic assessment and digitalized feedback to L2 learners' grammar acquisition: Technology-enhanced learning facilitation theory. *Journal of Educational Computing Research*, *63*(1), 147–172. https://doi.org/10.1177/07356331241296889

Lai, Z.CC. (2026). AI in EFL blended learning: impact on speaking performance and learning resilience. *Asian-Pacific Journal of Second and Foreign Language Education. 11*, 11. https://doi.org/10.1186/s40862-025-00367-4

Lantolf, J. P., & Poehner, M. E. (2014). *Sociocultural theory and the pedagogical imperative in L2 education*. Routledge.

Lantolf, J. P., & Thorne, S. L. (2006). *Sociocultural theory and the genesis of second language development*. Oxford University Press.

Lekamge, R., & Smith, C. (2025). Impact of auto-correction features in text-processing software on the academic writing of ESL learners. *ELOPE: English Language Overseas Perspectives and Enquiries*, *22*(1), 69–91. https://doi.org/10.4312/elope.22.1.69-91

Li, L., Zhang, X., Zou, B., & Yang, Q. (2025). AI partner or peer partner? Exploring AI-mediated interaction in EFL pronunciation from a socio-cultural perspective. *Learning, Culture and Social Interaction*, *55*, Article 100958. https://doi.org/10.1016/j.lcsi.2025.100958

Macías-Borrego, M. (2026). Artificial intelligence and error analysis: Effects on feedback of recurrent errors and fossilisation tendencies. *Education Sciences*, *16*(3), Article 393. https://doi.org/10.3390/educsci16030393

Mirza Suzani, S. (2025). A comparison of Iranian EFL learners' spelling knowledge enhancement through computer-mediated vs. conventional written corrective feedback. *Applications of Language Studies*, e720284. https://doi.org/10.22034/jals.2025.2042702.1046

Mooney, G., & Imandin, L. (2026). Leveraging artificial intelligence in higher education: A Vygotsky-inspired framework. *African Perspectives of Research in Teaching and Learning*, *10*(1), 184–197. https://doi.org/10.70875/v10i1article14

Murtisari, E. T., Januardi, J. I., Bonar, G., & Kurniawan, D. (2025). Beyond error correction: Lower-proficiency EFL learners' engagement with Grammarly. *Language Awareness*, advance online publication. https://doi.org/10.1080/09658416.2025.2585004

Nabilou, M., Torki, R., & Moradi, P. (2019). A microgenetic development study of learning English grammar through CALL in Iranian EFL beginner learners. *Journal of Foreign Language Research*, *9*(3), 763–787. https://doi.org/10.22059/jflr.2019.264666.547

Nabilou, M., & Zarei, A. A. (2023). The effects of three types of flipped learning through Shad application on EFL learners' lexical and grammatical achievement. *Journal of Modern Research in English Language Studies*, *10*(4), 113–133. https://doi.org/10.30479/jmrels.2023.18868.2219

Nabilou, M., & Zarei, A. A. (2025). Flipped instruction and its potential to control foreign language classroom anxiety through Shad application. *Journal of Applied Linguistics and Applied Literature: Dynamics and Advances*, *13*(1), 1–19. https://doi.org/10.22049/JALDA.2024.29687.1681

Nassaji, H., & Fotos, S. (2011). *Teaching grammar in second language classrooms: Integrating form-focused instruction in communicative context*. Routledge. https://doi.org/10.4324/9780203850961

Pust, D. (2026). Learning grammar the explorative way: Integrating interactive grammar animations into CALL. *Language Learning & Technology*, *30*(1), 1–21. https://doi.org/10.64152/10125/73667

Richards, J. C., & Rodgers, T. S. (2018). *Approaches and methods in language teaching* (3rd ed.). Cambridge University Press.

Salehi, D. H., Amiri, B. (2019). Impacts of using Microsoft Word (MS) software on Iranian EFL lecturers' grammar knowledge. *IJREE*. *4*(1), 1-10. https://doi.org/10.29252/ijree.4.1.1

Shavinka, D., & Retnowati, N. (2026). EFL students' experiences using Grammarly in English writing classrooms. In *The 6th Bogor English Student and Teacher (BEST) Conference 2026* (p. 49). Universitas Ibn Khaldun Bogor.

Sigvaldsen, N., & Kristiansen, K. (2024). *The effect of spell and grammar checkers on pupils' English: A study of texts written by Norwegian school children aged 13–14 years* [Master's thesis, UiT, The Arctic University of Norway]. Munin. https://munin.uit.no/handle/10037/33999

Simatupang, M. S., Tambunsaribu, G., Peter, R., Gunawan, R., & Murniarti, E. (2026). The mediating role of social influence in ChatGPT-assisted English learning performance: A sociocultural analysis in Indonesian higher education. *International Journal of Education and Practice, 14*(1), 554–579. https://doi.org/10.18488/61.v14i1.4770

Song, C. & Song, Y. (2023). Enhancing academic writing skills and motivation: assessing the efficacy of ChatGPT in AI-assisted language learning for EFL students. *Front. Psychol*. *14*:1260843. https://doi.org/10.3389/fpsyg.2023.1260843

Vygotsky, L. S. (1978). *Mind in society: The development of higher psychological processes*. Harvard University Press.

Wood, D., Bruner, J. S., & Ross, G. (1976). The role of tutoring in problem solving. *Journal of Child Psychology and Psychiatry*, *17*(2), 89–100. https://doi.org/10.1111/j.1469-7610.1976.tb00381.x

Yokokura, Y. (2026). AI-assisted self-paced grammar instruction in secondary education: Balancing individualized and collaborative EFL learning. *The English Teacher*, *54*(3), 1–18. https://doi.org/10.52696/LZVK1985

Zhang, C. (2026). Unlocking academic gains in smart-classroom settings: A moderated-mediation study among Chinese undergraduates. *Acta Psychologica*, *262*, Article 106175. https://doi.org/10.1016/j.actpsy.2026.106175

Zhang, R., & Zou, D. (2022). Types, purposes, and effectiveness of state-of-the-art technologies for second and foreign language learning. *Computer Assisted Language Learning, 35*(4), 696–742. https://doi.org/10.1080/09588221.2020.1744666