EFL Writing using AI Feedback: A Study of Effects on Syntax and Lexis

Arif Ahmed Mohammed Hassan Al-Ahdal[1], Abdullah Ali Altamimi[2]

[1]Department of English Language and Literature, College of Languages and Humanities, Qassim University, Buraydah, Saudi Arabia. Email: aa.alahdal@qu.edu.sa, ORCID ID: https://orcid.org/0000-0002-6747-0897

[2]Department of Languages and Translation, University College of Duba, University of Tabuk, Saudi Arabia. Email: aaltamimi@ut.edu.sa ,ORCID ID:https://orcid.org/0000-0002-7697-0535

HOW TO CITE:
Arif Ahmed Mohammed Hassan Al-Ahdal, Abdullah Ali Altamimi (2026). EFL Writing using AI Feedback: A study of effects on Syntax and Lexis. International Journal of Special Education, 41(10s), 712-724

COPYRIGHT STATEMENT:
Copyright: © 2026 Authors.
Open access publication underthe terms and conditions of the Creative CommonsAttribution (CC BY) license (http://creativecommons.org/licenses/by/4.0/).

ABSTRACT:
This study examines the effectiveness of AI-powered adaptive feedback in learning and its impact on improving the syntactic and lexical sophistication of tertiary English as a Foreign Language (EFL) learners. Although technology-enhanced language learning (TELL) classrooms have flourished immensely in recent times, there is little empirical evidence on the differential effects of computer-based intelligent feedback on language production. Drawing on the principles of Sociocultural Theory and Interaction Hypothesis, this research adopted a quasi-experimental approach with a sample of 90 undergraduate EFL students who were randomly allocated to AI-assisted and traditional feedback groups for a 12-week teaching cycle. Computational linguistic measures (Coh-Metrix and the Tool for the Automatic Analysis of Lexical Sophistication (TAALES)) were employed to collect data on syntactic and lexical indicators from the pre- and post-test phases. Results indicated statistically significant improvements among participants in the AI-assisted feedback condition in clause density, subordination index, and type-token ratio over those who received conventional feedback. The study concluded with the prediction that AI-induced metalinguistic awareness is a key mediator in the writing process and should be adopted in EFL classrooms. The study results have profound implications for curriculum designers, EFL teachers, and educational technologists exploring the use of intelligent tutoring systems (ITS) in EFL writing analysis and pedagogies, especially in low-resource educational institutions in Saudi Arabia.

Keywords: Adaptive learning systems, AI-powered feedback, EFL writing, lexical diversity, research work, syntactic complexity.

Introduction

The surge in integration of artificial intelligence (AI) tools in conventional classrooms is reshaping pedagogical approaches to teaching language(s). AI-based Adaptive Feedback Systems mark a paradigm shift in EFL writing instruction which has traditionally relied on a feedback-intensive, one-size-fits-all approach, and could have transformative impact on language learning results. Singh et al. (2025) defined AI tools "as an explanation for the dependent variable learning performance. AI

tools are technologies used in higher education to facilitate learning and instruction" (p.4).

Writing is broadly recognized as a cognitively challenging skill in L2 acquisition due to the many processes that it requires to operate simultaneously such as the coordination of syntactic processing, lexical retrieval, genre awareness, and pragmatic competence (Crossley, 2020). Current feedback methods are helpful but have several challenges such as, temporal, logistical, and expertise related limitations that are not easily addressed by AI technologies. The rise of automated writing evaluation (AWE) systems and natural language processing (NLP)-based feedback tools has enabled immediate, individual, and diagnostically accurate feedback on complex linguistic features (Barrot, 2021).

AI-based feedback can have greater micro reach than traditional instructor feedback which tends to be holistic, delayed, and variable in quality. Moreover, AI-based feedback can be offered within sentences, with patterns at smaller levels of detail that are syntactically related to a given sentence and offer richer lexical assessments and explanations in tandem with individual learner's metalinguistic profile orientation. Importantly, the affordances are very similar to those most recently identified in contemporary second language acquisition (SLA) theories as optimal conditions for second language acquisition: noticing, form-function mapping, and iterative reformulation (Delgado-Garza & García Mayo, 2025).

Literature Review

Literature now supports the potential of automated feedback tools to encourage deliberate engagement with form in the development of L2 writing (Barrot, 2022; Ranalli, 2021; Al-Ahdal et al., 2014; Al-Ahdal & Hameed, 2025).

Although there is increasing interest in what AI tools can add to language teaching, empirical evidence based on the specific impacts of AI tools on measurable language dimensions is still patchily developed and not as methodologically consistent as might be desired. However, research has shown generally positive relationships between IWF and overall writing quality (Guo et al., 2022; Wei et al., 2023; Aljabr & Al-Ahdal, 2024; Alkodimi & Al-Ahdal, 2021) but there are limited studies investigating the effect of IWF on specific structural and lexical parameters using fine-grained criteria. The evidence-in-practice results from Zhai and Ma (2023) which pooled the results of 26 primary studies, support a sizable overall effect of AWE ($g = 0.861$) on writing quality but with significant variation in results for specific linguistic outcome measures. This diversity highlights the importance of investigating each measure at a micro level.

In the background of this corpus, this study is based on two parallel theoretical frameworks: the AI tools feedback mechanism is regarded as a semiotic mediation tool within Vygotsky's (1978) Sociocultural Theory which defines language learning as a mostly sociocognitive process that relies on suitable tools and social interactions, enabling learners to gain deeper understanding of linguistic norms that would not be possible on their own (Xu & Zhang, 2022). At the same time, it is recognized that focusing attention on non-target-like forms is at the center of Long's (1996) Interaction Hypothesis which emphasizes the importance of negotiated interaction and corrective feedback. Both feedback processes are made operational in the form of structured metalinguistic commentary and deliberate revision tasks and are made scalable in a digital environment (Delgado-Garza & García Mayo, 2025).

In this study, two value-added analytical tools, Coh-Metrix (McNamara et al., 2014), and TAALES (Kyle et al., 2018), provided a rare level of precision and repeatability. Coh-Metrix produces more than 100 indexes of syntactic complexity, text cohesion, and readability, and TAALES offers a wealth of indexes of lexical sophistication such as frequency profiles, contextual diversity, and range. The measure of textual lexical diversity (MTLD) was also used with the Tool for the Automatic Analysis of Lexical Diversity (TAALED) (Kyle et al., 2021)

which addresses common problems with a simple type-token ratio for length correction. These instruments allow for a cross, theoretically driven, and empirically tractable multidimensional characterization of L2 writing development (Lu, 2010; Crossley & McNamara, 2014).

Previous AWE studies have shown that with certain precautions, automated feedback can encourage syntactic complexity. Lu (2011) showed that syntactic complexity measures serve as reliable indicators of writer development in ESL, and Lei et al. (2023) found that both length-based and phrasal syntactic complexity measures are reliable for measuring ESL writing development in longitudinal studies across four semesters. Concerning lexis, Kyle and Crossley (2015) confirmed that the TAALES indices are highly reliable predictors of lexical proficiency, and Crossley et al. (2016) found that taxonomic aspects such as cohesive and lexical characteristics gradually emerge in L2 academic writing. This study extends directly upon these computational linguistic rules and premises and introduces AI feedback as a pedagogical tool and addresses three interrelated concerns. First, whether AI feedback leads to considerable improvements in syntactic complexity (clause density, subordination index, and mean length of T-unit) during a 12-week intervention period compared to traditional teacher feedback. Second, whether there is significant difference between lexical diversity (indices TTR and MTLD) in the experimental group that receives feedback in an AI-assisted modality compared to the control group that receives feedback in a traditional modality. Third, if metalinguistic awareness is a statistically significant mediator between AI feedback engagement and any measure of writing quality outcomes. In addition to addressing these concerns, this research helps build empirical SLA research data and pedagogical perspectives for EFL practitioners engaged in digitally changing academic settings.

Materials and Methods

A quasi-experimental design with pre-post-tests administered to two groups was adopted in this study. The study was conducted at a major public university in the Western Province of Saudi Arabia over a full semester (12 weeks of instruction). Approval was obtained from the host institution's Institutional Review Board before data collection began, and written informed consent was obtained from all participating ninety undergraduate students enrolled in an intermediate academic writing course for EFL. Comparability of the experimental and control groups at baseline was achieved by stratification by gender and CEFR proficiency level. All participants were Saudi nationals studying English as a foreign language with Arabic as the mother tongue. Proficiency levels of all participants ranged from B1 to C1 level on the Common European Framework of Reference. All linguistic pre-test equivalence independent-samples t-tests showed that there were no initial differences between the groups for all variables (all ps> .20), indicating comparability between the groups at baseline. The ages, gender, and employment status of the sample participants in this study are presented in Table 1.

Table 1 Demographic Profiles of the participants

Variable	Category	AI Group (n = 45)	Control Group (n = 45)	Total (N = 90)
Gender	Male	24 (53.3%)	23 (51.1%)	47 (52.2%)
	Female	21 (46.7%)	22 (48.9%)	43 (47.8%)
Age (years)	18–20	18 (40.0%)	19 (42.2%)	37 (41.1%)
	21–23	22 (48.9%)	21 (46.7%)	43 (47.8%)
	24+	5 (11.1%)	5 (11.1%)	10 (11.1%)

Variable	Category	AI Group (n = 45)	Control Group (n = 45)	Total (N = 90)
Academic major	English Language	19 (42.2%)	20 (44.4%)	39 (43.3%)
	Education	14 (31.1%)	13 (28.9%)	27 (30.0%)
	Other	12 (26.7%)	12 (26.7%)	24 (26.7%)
CEFR Proficiency	B1	11 (24.4%)	12 (26.7%)	23 (25.6%)
	B2	26 (57.8%)	25 (55.6%)	51 (56.7%)
	C1	8 (17.8%)	8 (17.8%)	16 (17.8%)

Note. CEFR = Common European Framework of Reference. Percentages may not sum to 100 due to rounding.

The instructional intervention lasted 12 sessions during which both groups completed four identical genre-specific essay writing tasks (argumentative, expository, compare-and-contrast, and cause-and-effect). The writing tasks involved process-oriented approach including pre-writing (brainstorming), drafting, peer revision, and revision. The first important difference between the conditions was in the way feedback was shared in the revision process. The two groups were given three hours of equal face-to-face classroom instruction per week to develop rhetorical strategies and writing process skills to isolate feedback modality as the independent variable. The AI-assisted group engaged with an adaptive feedback tool and an NLP engine to analyze the corpus. The system provided feedback on syntactic structures such as clause embedding, subordination types, sentence length variation, and academic vocabulary deployment/exposure, collocational appropriateness, and cohesion in the overall text. Importantly, metalinguistic explanations which were rule-suggestive and explanatory, attracted learners' attention to the structural and functional basis of flagged errors as well as possible improvements that were incorporated into the system (Barrot, 2021; Delgado-Garza & García Mayo, 2025). Feedback was provided to learners asynchronously on a dedicated LMS portal with at least two revision cycles for each essay and documentation of their revisions was done in a reflective revision log. Platform usage logs were examined to ensure usage of the platform and sufficient engagement (no less than 30 minutes per feedback session). Finally, the control group was offered written corrective feedback by the course teacher, a trained TESOL expert with eight years of experience teaching EFL writing. The instructor's comments were similar to conventional ones found in the AWE literature (Ajabshir & Ebadi, 2023; Thi & Nikolov, 2023) indicating global coherence, thesis development, and grammatical accuracy throughout the text. This feedback was offered within 72 hours of submission of the essays, as was the case for the learner's final revision.

The syntactic complexity indices in this study were derived from learners' texts using Coh-Metrix version 3.0 (McNamara et al., 2014) which calculates indices such as clause density, subordination index, and mean length of T-unit (MLT) broadly validated as indicators for the L2 syntactic level (Crossley & McNamara, 2014; Lu, 2010, 2011; Lei et al., 2023). TAALES version 2.2 (Kyle et al., 2018) was used to extract lexical sophistication measures for the texts. Lexical diversity was also examined across datasets using TAALED (Kyle et al., 2021) and the MTLD index, a text-length corrected index of lexical diversity. All texts in the pre- and post-tests were so processed and analyzed without identifying the writer. To operationalize metalinguistic awareness, two linguistic indicators were operationalized: awareness of grammatical structures and awareness of words/expressions and feedback utilization, (Barac et al., 2014; Gunnerud, 2021) as well as

two behavioral indicators, revision behavior and vocabulary knowledge with the help of a self-report instrument that was adapted to accommodate the 20 items used in the study (1 = definitely not, 5 = definitely yes).

In the current sample, the instrument was found to be reliable with a Cronbach's α of 0.87 for the Experimental Group and 0.84 for the Control Group. The post-test was administered after the last written task in Week 12. Data was analyzed statistically using the R software package version 4.3.2, the lavaan, and lme4 packages. The pre- and post-test outputs were analyzed using ANOVA with between-subjects factor (AI vs control) applied by time (pre- vs post-test) tested within-subjects factor. Partial eta squared (η2) was used to measure the effect sizes. To test the hypothesized mediating role of metalinguistic awareness, Baron and Kenny's (1986) steps of mediation procedure were used along with the PROCESS macro (Hayes, 2018, 5,000 iterations) to generate bootstrap confidence intervals for the mediating model. All tests were set for a significance level of α = .05.

Results and Discussion

With baseline equivalence theory for significant intergroup differences on all the linguistic features (clause density: t(88) = 0.67, p = .50; TTR: t(88) = 0.42, p = .68; MTLD: t(88) = 0.31, p = .76), the credibility of post-test comparisons was maintained. All primary linguistic outcomes are presented as descriptive statistics and in terms of the results of mixed design-ANOVA analyses in Table 2.

Table 2 Descriptive Statistics and ANOVA results for linguistic outcomes (pre- and post tests)

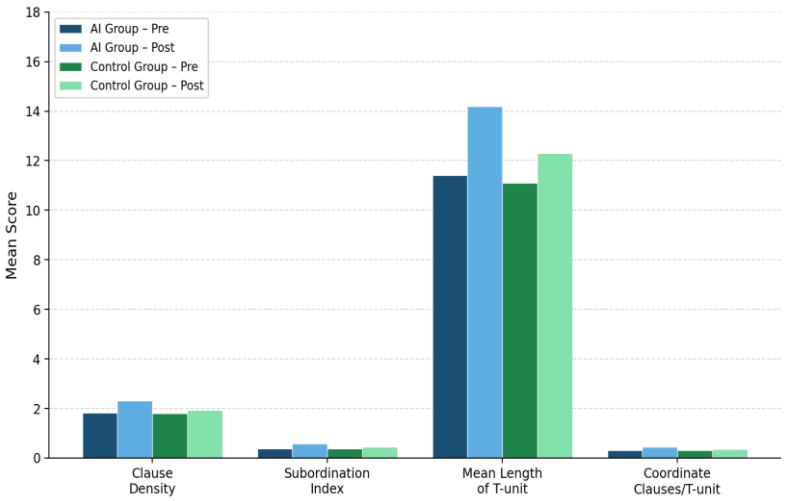
Measure	AI Pre	AI Post	Ctrl Pre	Ctrl Post	F	p	η^2
Clause Density	1.82 (0.21)	2.31 (0.19)	1.79 (0.22)	1.94 (0.20)	18.43	.000**	.29
Subordination Index	0.38 (0.07)	0.57 (0.08)	0.37 (0.06)	0.43 (0.07)	22.17	.000**	.33
Mean Length of T-unit	11.4 (1.8)	14.2 (1.6)	11.1 (1.9)	12.3 (1.7)	15.62	.000**	.26
Type-Token Ratio	0.412 (0.04)	0.541 (0.05)	0.408 (0.04)	0.451 (0.05)	24.89	.000**	.37
MTLD Score	62.3 (8.1)	81.7 (7.4)	61.9 (8.3)	70.2 (7.8)	19.34	.000**	.30
Lexical Sophistication	2.31 (0.31)	3.12 (0.28)	2.29 (0.30)	2.67 (0.29)	16.77	.000**	.27
Metalinguistic Awareness	2.74 (0.45)	4.18 (0.39)	2.71 (0.43)	3.19 (0.41)	28.54	.000**	.40

Note. Values represent M (SD). F-ratios reflect Group × Time interaction effects. η^2 = partial eta squared. * p < .05. ** p < .001.

The most significant result from the syntactic complexity analyses was the significant Group × Time interaction for clause density (F(1, 88) = 18.43, p < .001, η^2 = .29), with the Experimental Group (AI group) showing significantly higher gains in clause density between the pre- and post-test (Mpre = 1.82, SD = 0.21; Mpost = 2.31, SD = 0.19) than the Control Group (Mpre = 1.79, SD = 0.22; Mpost = 1.94, SD = 0.20). A similar effect was observed with the subordination index (F(1, 88) = 22.17, p < .001, η^2 = .33), which increased from a pre-test mean

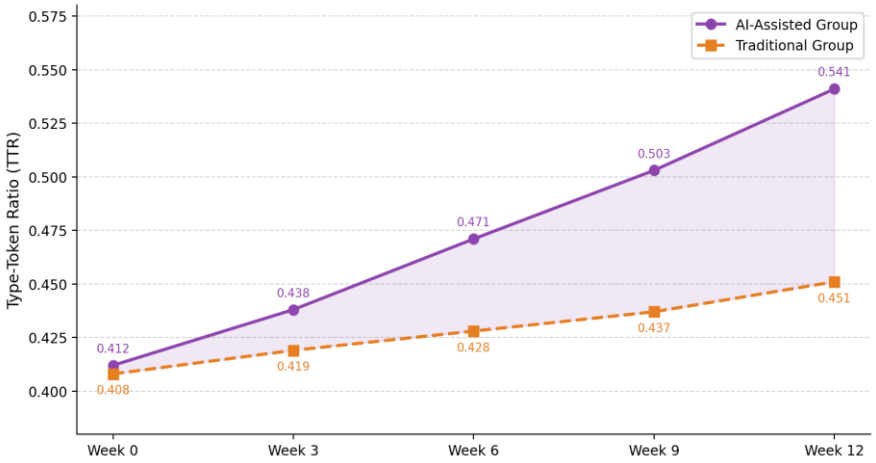
of 0.38 to a post-test mean of 0.57 for the AI group. The mean length of the T-unit similarly showed a significant interaction ($F(1, 88) = 15.62$, $p < .001$, $\eta^2 = .26$). Figure 1 shows the syntactic complexity results.

Figure 1. Pre- and Post-Test Syntactic Complexity Measures by Group



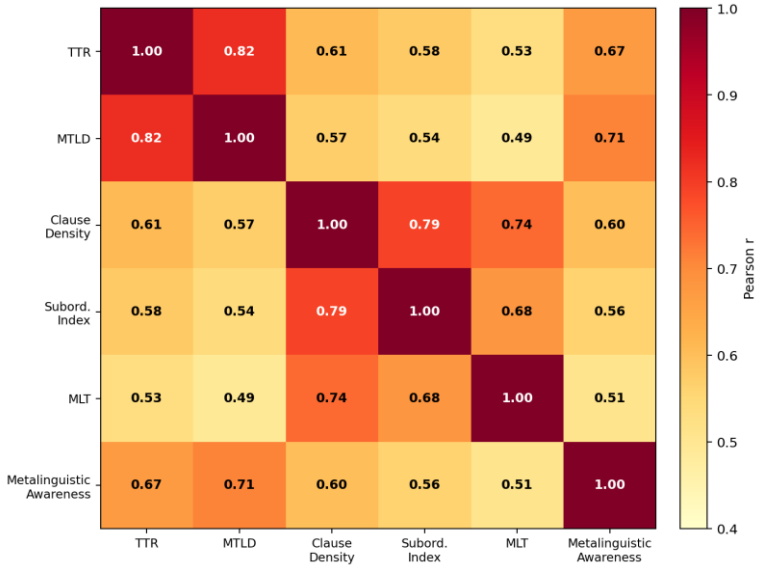
These results corroborate and confirm previous studies on AWE and syntactic complexity. Despite the limited evidence provided by Thi and Nikolov (2023) in a 13-week study for both the teacher and automated feedback groups on syntactic complexity (L2SCA) applied to their indices, the current study shows that a specialized AI adaptive feedback system along with metalinguistic commentaries can yield notable gains on syntactic complexity. Differences can be explained by the fact that the form-focused and targeted AI platform used here focused on a specific form for specific definition levels contrasting with a less specific form system used in other settings that provided only code-switching feedback to assist the adult L2s in noticing. The differences may be due to the fact that the platform used in this intervention was a focused, specific AI tool that went beyond providing generic corrective feedback which could be an AWE, but instead provided clause-level explanations of the structures of the English language that were in line with metalinguistic feedback found to facilitate noticing in adult L2 learners (Delgado-Garza & García Mayo, 2025). These findings are consistent with the results of Ajabshir and Ebadi (2023) who showed that AWE yielded higher scores than teacher-focused feedback in syntactic complexity for the two genres of narrative and argumentative texts. The longitudinal pathways of TTR for the two groups over the course of the 12-week intervention are shown in Figure 2 indicating a greater and sharper growth trajectory in the AI group compared to the Control Group, beginning at Week 3. The post-test compared to baseline TTR showed a significant improvement for the Experimental Group (+0.129 points) and a less significant improvement for the Control Group (+0.043 points). The largest effect size was for the differential, with a level of .37 ($F(1, 88) = 24.89$, $p < .001$). The MTLD results similarly showed a significant Group × Time interaction ($F(1, 88) = 19.34$, $p < .001$, $\eta^2 = .30$).

Figure 2. Longitudinal Trajectory of Lexical Diversity (TTR) Across 12 Weeks



The lexical diversity score achieved in the Experimental Group might be explained by the platform's ability to emphasize the importance of enriching vocabulary at the writing stage, offer lexical alternatives, provide collocational ideas, and suggest register-appropriate vocabulary on a highly academic level. This is also in line with the results obtained by Kyle et al. (2021) who showed that MTLD is sensitive to instructed target word construction. It is particularly important to note the superior MTLD gains in the AI tools condition because MTLD is relatively unaffected by text-length artifacts, as was simple TTR (Kyle et al., 2021), enhancing confidence in the reliability of this gain. Dizon and Gayed (2021) also examined lexical richness in L2 contexts, using Grammarly as an AWE program. They found lexical benefits, a finding which supports the current results. A correlation heatmap of all post-test language variables for the Experimental Group is shown in Figure 3. A complex pattern of interrelationships between the syntactic and lexical scores is apparent from the heatmap, particularly with regard to the strong relationships between TTR and MTLD (r = .82), clause density and subordination index (r = .79), and metalinguistic awareness scores with both lexical (r = .71) and syntactic (r = .60) awareness. The convergent validity of TTR and MTLD represents results similar to those of Kyle et al. (2021), and metalinguistic awareness and its relationship with linguistic outcomes emphasize the need for mediation analysis to be conducted.

Figure 3. Correlation Heatmap of Key Linguistic Variables (Post-Test, AI Group)



Summary of the mediation analyses in Table 3 supports the hypothesized mediation of metalinguistic awareness. Results indicated that the indirect effect of AI feedback on syntactic complexity via metalinguistic awareness (β = 0.33, 95% CI [0.19, 0.50]) and on lexical diversity was significant (β = 0.28, 95% CI [0.15, 0.43]). Whether mediated by syntactic complexity (β = 0.29, p = .010) or lexical diversity (β = 0.31, p = .012), these two direct effects persisted after controlling for these two mediators when statistically evaluated, indicating partial mediation. This is in line with the theoretical construct such that metalinguistic awareness is also a meaningful mechanism, and more direct effects are likely due to higher rates of revision throughout the writing process, better student motivation, and decreased anxiety about processing (Barrot, 2021; Ranalli, 2021).

Table 3 Mediation Analysis: Metalinguistic Awareness as Mediator Between AI Feedback and Linguistic Outcomes

Pathway	β	SE	95% CI	t	p
AI Feedback → Metalinguistic Awareness (a path)	0.61	0.09	[0.43, 0.79]	6.78	.000**
Metalinguistic Awareness → Syntactic Complexity (b path)	0.54	0.10	[0.34, 0.74]	5.40	.000**
AI Feedback → Syntactic Complexity (direct, c' path)	0.29	0.11	[0.07, 0.51]	2.63	.010*
AI Feedback → Syntactic Complexity (total, c path)	0.62	0.10	[0.42, 0.82]	6.20	.000**
Indirect Effect via Metalinguistic Awareness (a×b)	0.33	0.08	[0.19, 0.50]	—	—
AI Feedback → Lexical Diversity (direct, c' path)	0.31	0.12	[0.07, 0.55]	2.58	.012*
Indirect Effect via Metalinguistic Awareness	0.28	0.07	[0.15, 0.43]	—	—

Note. Bootstrap confidence intervals based on 5,000 resamples. β = standardized regression coefficient. CI = confidence interval. * p < .05. ** p < .001.

Results of the post-test metalinguistic awareness survey are displayed in Figure 4. Most of the participants in the Experimental Group evidenced improved awareness of sentence complexity (78%), self-assessment of vocabulary range (71%), and use of revision strategies (73%). The AI feedback system used in this study not only brought about quantitative linguistic improvement, but also, as reflected in the attitudinal changes, deepened the students' metacognitive awareness of the writing process, which echoes Jiang and Yu's (2022) findings with Chinese EFL students' metacognitive awareness of the writing process. In a mixed methods study comparing teachers' and automated feedback, Wang and Han (2022) found that automated feedback had a positive effect on cognitive processing of writing quality.

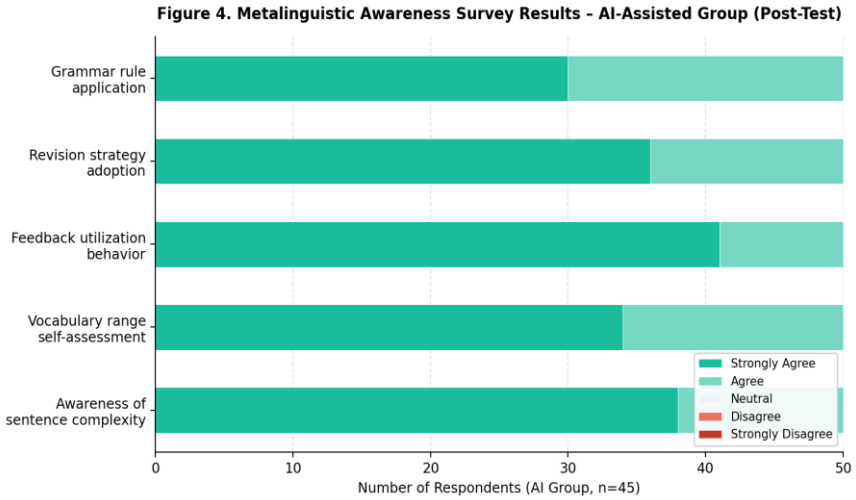
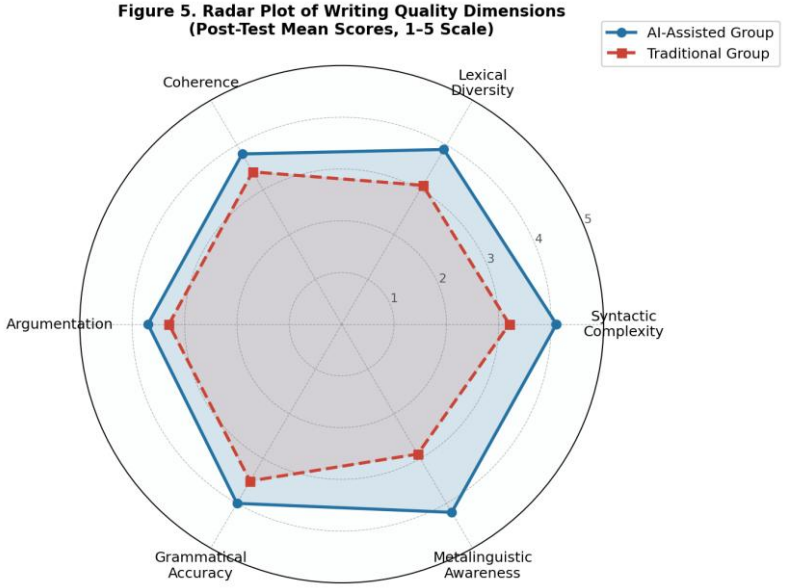


Figure 5 illustrates a multidimensional picture of writing quality in both groups by plotting them across six dimensions on a radar graph. Generally, the Experimental Group showed higher scores than the Control Group in all three areas with the differences being greatest in metalinguistic awareness (AI: 4.2 vs. control: 2.9), syntactic complexity (AI: 4.1 vs. control: 3.2), and lexical diversity (AI: 3.9 vs. control: 3.1). Interestingly, the Control Group exhibited a relatively low (3.4) coherence score on the dimensions, implying that the traditional feedback provided by an instructor can be found to be more effective in the development of macro-level discourse organization. This finding is in line with Crossley et al.'s (2016) analysis of cohesion development in L2 writing and Mao et al.'s (2024) synthesis of naturalistic classroom feedback studies.



The present results (η^2 = 0.26-0.40) are similar to or better than those reported by the meta-analysis conducted by Zhai and Ma (2023) which showed a moderately large positive mean effect overall (η^2 = 0.861) for AWE on writing quality across 26 studies, and that the effect of AWE in a natural language setting was more positive for EFL/ESL learners than for native English learners. As opposed to the simple documentation of effects, this partial mediation structure is an explanatory modeling step that identifies metalinguistic awareness as a concrete psychological mechanism that transduces the effects of AI feedback. Combining quantitative linguistic metrics, self-report attitudinal measures, and structural equation models

provides a plurality of methodological means for validating the effectiveness of this intervention (Li et al., 2015; Crossley, 2013).

The study's findings have implications for teaching practice in EFL settings in Saudi Arabia. The limited resources in EFL classrooms, coupled with large class size, low contact hours, and the different levels of knowledge of the teachers, have historically restricted the scope for providing instant feedback. Automated AI adaptive feedback systems could help to make high-quality writing instruction accessible to all in this context. The reported benefits of AI mark-making feedback over human teaching feedback for syntactic complexity and lexical diversity, while maintaining some benefits in respect to coherence, indicate that a blended approach could be pedagogically the most effective: provided by AI formative feedback for syntactic complexity and lexical diversity, but supplemented by human teaching for discourse-level organization (Ajabshir & Ebadi, 2023; Wei et al., 2023).

Conclusion

Findings of this study offer insights into the feasibility of using AI adaptive feedback systems to boost EFL learners' writing proficiency, especially syntactic complexity and lexical range of their writing. The quasi-experimental design results showed statistically significant differences in the performance of the two groups in terms of key linguistic metrics. These results highlight the importance of AI tools in providing timely, targeted, and comprehensive feedback that complies with current SLA theories. Additionally, the mediational nature of metalinguistic awareness created by AI tools in the writing process has great implications for both teachers and curriculum developers. It gives learners more opportunities to interact with the complexity of the language, enabling them to write with greater fluency and better understanding of form and function of language. This research highlights the value of using IMS in ELT environments, particularly in the context of resource-constrained instruction, in Saudi Arabia and other non-English speaking countries. With the potential of AI-driven feedback systems demonstrated, there is an opportunity for continued research to investigate the long-term effects of AI tools in language learning and consider the scalability of the systems in a variety of language learning settings. The results call for a revolutionary approach to teaching and learning of EFL writing in the future. Not only is AI a tool for EFL writing classes, it is also a significant companion in enhancing language learning outcomes at the same time.

Recommendations

Based on the findings, this study recommends enrolling large samples across several institutional settings in future studies. Replication in a variety of educational settings is strongly recommended, as the single-institution design restricts the generalizability of the results to other Saudi and/or Gulf EFL contexts. In addition, developing and evaluating culturally adapted AI feedback tools is a promising avenue for both applied linguistics research and educational technology design.

Future research could benefit from adding stimulated recall sessions and/or self-report measures to think aloud or behavioral trace data from learner revision logs. These other data sources would ecologically validate the metalinguistic processing mechanisms found to be central to the mediation structure reported here.

Variables such as anxiety in learning, intrinsic motivation, digital knowledge and exposure, and working memory should be studied in following studies.

Limitations

The quasi-experimental design in this study might have led to group selection bias or differential motivation among groups. Moreover, the L2FLP used was not tailored to the Saudi Arabian program of English as a foreign language (EFL) or to Arabic–English interlanguage patterns. Longitudinal studies following students through the breadth of a full

academic year or over a full school year would be more informative with respect to durability and cumulative effects, including those of the embeddings of the complex syntactic pattern. In addition, metalinguistic awareness was measured exclusively by self-report methods which may have caused some bias.

References

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), Article 26. https://doi.org/10.1186/s40862-023-00201-9

Al-Ahdal, A. A. M. H., Alfallaj, F. S., Al-Awaied, S. A., & Al-Hattami, A. A. (2014). A comparative study of proficiency in speaking and writing among EFL learners in Saudi Arabia. American International Journal of Contemporary Research, 4(2), 141–149.

Al-Ahdal, A.A.M.H.., & Hameed, P.F.M., (2025). The Contribution of Writing Portfolios in Developing EFL Students' Writing Accuracy and Fluency: An Exploratory Study. Forum for Linguistic Studies. 7(10): 708–724. https://doi.org/10.30564/fls.v7i10.10942

Aljabr, F. S., & Al-Ahdal, A. A. M. H. (2024). Ethical and pedagogical implications of AI in language education: An empirical study at Ha'il University. Acta Psychologica, 251, https://doi.org/10.1016/j.actpsy.2024.104605 .

Alkodimi, K. A., & Al-Ahdal, A. A. M. H. (2021). Strategies of teaching writing at Saudi tertiary-level institutions: Reality and expectations. Arab World English Journal, 12(2), 399-413.

Barac, R., Bialystok, E., Castro, D. C., & Sanchez, M. (2014). The cognitive development of young dual language learners: A critical review. Early childhood research quarterly, 29(4), 699-714. https://doi.org/10.31265/usps.72

Baron, R. M., & Kenny, D. A. (1986). The moderator–mediator variable distinction in social psychological research: Conceptual, strategic, and statistical considerations. Journal of Personality and Social Psychology, 51(6), 1173–1182. https://doi.org/10.1037/0022-3514.51.6.1173

Barrot, J. S. (2021). Using automated written corrective feedback in the writing classrooms: Effects on L2 writing accuracy. Computer Assisted Language Learning, 36(4), 584–607. https://doi.org/10.1080/09588221.2021.1936071

Barrot, J. S. (2022). Integrating technology into ESL/EFL writing through Grammarly. RELC Journal, 53(3), 764–768. https://doi.org/10.1177/0033688220966632

Crossley, S. A. (2013). Advancing research in second language writing through computational tools and machine learning techniques: A research agenda. Language Teaching, 46(2), 256–271. https://doi.org/10.1017/S0261444812000560

Crossley, S. A. (2020). Linguistic features in writing quality and development: An overview. Journal of Writing Research, 11(3), 415–443. https://doi.org/10.17239/jowr-2020.11.03.01

Crossley, S. A., & McNamara, D. S. (2014). Does writing development equal writing quality? A computational investigation of syntactic complexity in L2 learners. Journal of Second Language Writing, 26, 66–79. https://doi.org/10.1016/j.jslw.2014.09.006

Crossley, S. A., Kyle, K., & McNamara, D. S. (2016). The development and use of cohesive devices in L2 writing and their relations to judgments of essay quality. Journal of Second Language Writing, 32, 1–16. https://doi.org/10.1016/j.jslw.2016.01.003

Delgado-Garza, P., & García Mayo, M. D. P. (2025). Can we train young EFL learners to 'notice the gap'? Exploring the relationship between metalinguistic awareness, grammar learning and the use of metalinguistic explanations in a dictogloss task. International Review of Applied Linguistics in Language Teaching, 63(3), 1573-1597.https://doi.org/10.1007/978-3-319-24277-4_9

Dizon, G., & Gayed, J. M. (2021). Examining the impact of Grammarly on the quality of mobile L2 writing. JALT CALL Journal, 17(2), 74–92. https://doi.org/10.29140/jaltcall.v17n2.336

Gunnerud, H. L. (2021). Bilingualism: Advantages and Disadvantages in Cognitive Processing, Language and Reading Comprehension. https://doi.org/10.31265/usps.72

Guo, Q., Feng, R., & Hua, Y. (2022). How effectively can EFL students use automated written corrective feedback in research writing? Computer Assisted Language Learning, 35(9), 2312–2331. https://doi.org/10.1080/09588221.2021.1879161

Hayes, A. F. (2018). Partial, conditional, and moderated moderated mediation: Quantification, inference, and interpretation. Communication monographs, 85(1), 4-40.

Jiang, L., & Yu, S. (2022). Appropriating automated feedback in L2 writing: Experiences of Chinese EFL student writers. Computer Assisted Language Learning, 35(7), 1329–1353. https://doi.org/10.1080/09588221.2020.1815099

Kyle, K., & Crossley, S. A. (2015). Automatically assessing lexical sophistication: Indices, tools, findings, and application. TESOL Quarterly, 49(4), 757–786. https://doi.org/10.1002/tesq.194

Kyle, K., Crossley, S. A., & Berger, C. (2018). The tool for the automatic analysis of lexical sophistication (TAALES): Version 2.0. Behavior Research Methods, 50(3), 1030–1046. https://doi.org/10.3758/s13428-017-0924-4

Kyle, K., Crossley, S. A., & Jarvis, S. (2021). Assessing the validity of lexical diversity indices using direct judgements. Language Assessment Quarterly, 18(2), 154–170. https://doi.org/10.1080/15434303.2020.1844205

Kyle, K., Crossley, S. A., & Verspoor, M. (2021). Measuring longitudinal writing development using indices of syntactic complexity and sophistication. Studies in Second Language Acquisition, 43(4), 781–812. https://doi.org/10.1017/S0272263120000546

Lei, L., Wen, J., & Yang, X. (2023). A large-scale longitudinal study of syntactic complexity development in EFL writing: A mixed-effects model approach. Journal of Second Language Writing, 59, Article 100962. https://doi.org/10.1016/j.jslw.2022.100962

Li, J., Link, S., & Hegelheimer, V. (2015). Rethinking the role of automated writing evaluation (AWE) feedback in ESL writing instruction. Journal of Second Language Writing, 27, 1–18. https://doi.org/10.1016/j.jslw.2014.10.004

Lu, X. (2010). Automatic analysis of syntactic complexity in second language writing. International Journal of Corpus Linguistics, 15(4), 474–496. https://doi.org/10.1075/ijcl.15.4.02lu

Mao, Z., Lee, I., & Li, S. (2024). Written corrective feedback in second A corpus-based evaluation of syntactic complexity measures as indices of college-level ESL writers' language developmentlanguage writing: A synthesis of naturalistic classroom studies. Language Teaching, 57(4), 449–477. https://doi.org/10.1017/S0261444823000393

McNamara, D. S., Graesser, A. C., McCarthy, P. M., & Cai, Z. (2014). Automated evaluation of text and discourse with Coh-Metrix. Cambridge University Press. https://doi.org/10.1017/CBO9780511894664

Ranalli, J. (2021). Exploring L2 student engagement with automated written corrective feedback: Nature, variability and predictors. Computer Assisted Language Learning, 34(5–6), 597–624. https://doi.org/10.1080/09588221.2019.16174215

Singh, R., Singh, S. K., & Mishra, N. (2025). Factors influencing student learning performance and continuous use of artificial intelligence in online higher education. Discover Education, 4(1), 292. https://doi.org/10.1007/s44217-025-00728-8

Thi, N. K., & Nikolov, M. (2023). Effects of teacher, automated, and combined feedback on syntactic complexity in EFL students' writing. Asian-Pacific Journal of Second and Foreign Language Education, 8(1), Article 6. https://doi.org/10.1186/s40862-022-00182-1 22222222222222222222222222222222

Vygotsky, L. S. (1978). Socio-cultural theory. Mind in society, 6(3), 23-43.

Wang, Z., & Han, F. (2022). The effects of teacher feedback and automated feedback on cognitive and psychological aspects of foreign language writing: A mixed-methods research. Frontiers in Psychology, 13. https://doi.org/10.3389/fpsyg.2022.909802

Wei, P., Wang, X., & Dong, H. (2023). The impact of automated writing evaluation on second language writing skills of Chinese EFL learners: A randomized controlled trial. Frontiers in Psychology, 14, Article 1249991. https://doi.org/10.3389/fpsyg.2023.1249991

Xu, J., & Zhang, S. (2022). Understanding AWE feedback and English writing of learners with different proficiency levels in an EFL classroom: A sociocultural perspective. Asia-Pacific Education Researcher, 31(4), 357–367. https://doi.org/10.1007/s40299-021-00577-7

Zhai, N., & Ma, X. (2023). The effectiveness of automated writing evaluation on writing quality: A meta-analysis. Journal of Educational Computing Research, 61(4), 875–900. https://doi.org/10.1177/07356331221127300

Zhang, Z. V. (2020). Engaging with automated writing evaluation (AWE) feedback on L2 writing: Student perceptions and revisions. Assessing Writing, 43, Article 100439. https://doi.org/10.1016/j.asw.2019.100439