e-ISSN: 2528-8091
Vol. 11, N° 1 - 2026

Effectiveness of Automated Feedback Systems in Enhancing English Academic Writing Skills

Eficacia de los Sistemas de Retroalimentación Automatizada en la mejora de las habilidades de Escritura Académica en inglés

https://doi.org/10.5281/zenodo.19893794

Erika Paola Garcia León [1]
Universidad Técnica de Babahoyo, Ecuador
https://orcid.org/0000-0001-8203-6434
egarcia@utb.edu.ec

María Grazzia González Quinto [2]
Universidad Técnica de Babahoyo, Ecuador
https://orcid.org/0009-0003-3593-
mgonzalezq@utb.edu.ec

Pablo Luis Vásconez Mera [3]
Universidad Técnica de Babahoyo, Ecuador
https://orcid.org/0000-0001-9035-7166
pvasconezm@utb.edu.ec

Nelly Victoria Ley Leiva [4]
Universidad Técnica de Babahoyo, Ecuador
https://orcid.org/0000-0003-2296-7354
nley@utb.edu.ec

DIRECCIÓN PARA CORRESPONDENCIA: egarcia@utb.edu.ec
Fecha de recepción: 15/11/2025
Fecha de aceptación: 02/01/2026

ABSTRACT

The objective of this study was to analyze the impact of automated feedback tools on improving academic writing in English among university students, considering direct effects and explanatory mechanisms associated with mediating and moderating variables. A quantitative approach was applied with a quasi-experimental pretest-posttest design and control group with a population of 140 fifth-level students from the Language Center of the Technical University of Babahoyo, distributed in two parallel groups (control and experimental). The intervention integrated automated feedback platforms (Write &

Improve, Grammarly, Ginger, and Quillbot), and grammatical accuracy, coherence-cohesion, syntactic complexity, and lexical range were evaluated as dependent variables; intensity and type of feedback were evaluated as independent variables. The analyses included descriptive statistics, t-tests for related and independent samples, ANCOVA controlling for the pretest, and mediation models. The results described showed increases between the pretest and posttest, highlighting the intensity of feedback (M = 49.96 to 65.30) and grammatical accuracy (M = 69.76 to 79.39). In the experimental group, pre-post comparisons showed significant improvements in all dimensions ($p<0.001$), and the posttest comparison between groups confirmed significant advantages for the experimental group ($p<0.001$), with relevant mean differences in grammatical accuracy (MD = 18.07) and coherence-cohesion (MD = 17.38). The ANCOVA showed very significant adjusted differences by group when controlling for pretest, defending the effectiveness of the intervention. Additionally, mediation indicated that self-regulation partially explained the relationship between feedback and writing indicators (indirect = 0.049; 95% CI [0.013, 0.085]; p=.007). Taken together, these findings support the notion that automated feedback significantly improves the quality of English writing and that its impact is enhanced by self-regulation processes.

KEYWORDS: *Automated writing evaluation, Computer – mediated feedback, Formative feedback, Higher education contexts, Learner engagement.*

RESUMEN

El objetivo de este estudio fue analizar el impacto de herramientas de retroalimentación automatizada en la mejora de la escritura académica en inglés en estudiantes universitarios, considerando efectos directos y mecanismos explicativos asociados a variables mediadoras y moderadoras. Se aplicó un enfoque cuantitativo con diseño cuasiexperimental pretest – Postest y grupo de control con una población de 140 estudiantes de quinto nivel del Centro de Idiomas de la Universidad Técnica de Babahoyo, distribuidos en dos paralelos (control y experimental). La intervención integró plataformas de retroalimentación automatizadas (Write & Improve, Grammarly, Ginger y Quilbot) y se evaluaron como variables dependientes la precisión gramatical, coherencia - cohesión, complejidad sintáctica y rango léxico; como variables independientes, la intensidad y el tipo de retroalimentación. Los análisis incluyeron estadísticas descriptivas, pruebas t para muestras relacionadas e independientes,

e-ISSN: 2528-8091
Vol. 11, N° 1 - 2026

ANCOVA controlando el pretest y modelos de mediación. los resultados descritos mostraron aumentos entre pretest y Postest, destacando la intensidad de retroalimentación (M = 49,96 a 65,30) y la precisión gramatical (M = 69,76 a 79,39). En el grupo experimental las comparaciones pre – post mostraron mejoras significativas en todas las dimensiones (p<0,001) y la comparación Postest entre grupos confirma ventajas importantes del grupo experimental (p<0,001), con diferencias de medias relevantes en precisión gramatical (MD = 18,07) y coherencia – cohesión (MD = 17,38). El ANCOVA mostró diferencias ajustadas muy importantes por grupo al controlar pretest defendiendo la efectividad de la intervención. Adicionalmente, la mediación indicó que la autorregulación explicó parcialmente la relación entre retroalimentación e indicadores de escritura (indirect = 0,049; IC95% [0.013, 0.085]; p=.007). en conjunto, estos hallazgos sostienen que la retroalimentación automatizada mejora de manera significativa la calidad de la escritura del idioma inglés y que su impacto se potencia mediante procesos de autorregulación.

PALABRAS CLAVE: *Contextos de educación superior, Evaluación automatizada de escritura, Participación del alumno, Retroalimentación formativa, Retroalimentación mediada por ordenador.*

INTRODUCCIÓN

The integration of various automated feedback tools into artificial intelligence has transformed the process of teaching and assessing writing in English worldwide. These types of technologies allow for the analysis of grammatical, lexical, and discursive errors with high accuracy, which benefits more personalized learning. As a result, different types of educational systems are beginning to adopt these types of intelligent platforms in order to strengthen writing skills and respond to the communicative demands of globalized environments (Yang et al., 2024).

In Latin America, the frequent use of technologies for automated feedback has grown progressively, although there are still gaps in digital infrastructure, teacher training, and equitable access to advanced platforms. Nevertheless, many countries are promoting educational digitization policies aimed at improving the teaching of English as a foreign language. Thus, the application of intelligent tools becomes a viable alternative for

optimizing assessment processes and promoting autonomous learning (Rivadeneira et al., 2025).

In Ecuador, the incorporation of automated feedback systems is still in its infancy, despite progress in policies aimed at strengthening English language and digital education. Various higher education institutions have begun experimenting with AI-based platforms to improve academic writing in English; however, local empirical evidence remains limited. It is therefore important to analyze the real impact of these types of technologies in different educational contexts in Ecuador and their contribution to the development of language skills (Moreira et al., 2025).

This research is based on the hypothesis that automated feedback tools significantly influence the improvement of writing in English by offering immediate corrections and precise guidance. It justifies its relevance by providing evidence on the effectiveness of this type of technology in homogeneous contexts. In addition, it allows us to understand how feedback can intelligently complement teaching and promote more efficient, personalized learning practices geared toward student performance.

The overall objective of the following research is to analyze the impact of automated feedback tools on improving writing skills among university students. To this end, the following objectives are proposed: to determine students' perceptions regarding the use of these tools; to compare the performance of students who receive automated feedback with those who use traditional methods; and to identify the factors that facilitate or limit their implementation in education.

Fundamentals of Writing in English (L2 Writing)

The process of writing in English as a second language (L2) is a complex cognitive process that involves planning, textualization, and continuous revision of discourse. Various classical models, such as that of Flower and Hayes, argue that writing requires the simultaneous use of different linguistic systems, which represents a greater challenge for students. Writing proficiency includes grammatical accuracy, lexical variety, overall coherence, and textual cohesion, which are essential elements for producing clear and academically valid texts (Li et al., 2024).

Likewise, writing in L2 is influenced by different affective factors, such as levels of exposure to the English language and individual self-regulation strategies. Several recent studies highlight that student tend to transfer structures from their native language, which leads to persistent errors in syntax and textual organization. For this reason, various pedagogical approaches emphasize constant feedback and guided practice in order to develop autonomous writing oriented toward communicative performance (Tian et al., 2022).

Feedback in language teaching

The feedback process is a central component in language acquisition, as it identifies errors, adjusts linguistic production, and improves communicative accuracy. In the field of L2, corrective feedback can be direct, indirect, or metalinguistic, and its effectiveness depends on the type of error and the student's level. In addition, it promotes formative learning and provides actionable information that guides the student toward more appropriate performance (Amoli, 2020).

Similarly, the literature shows that feedback encourages reflection on language use, increases metalinguistic awareness, and promotes autonomous editing practices. Traditional strategies are often effective, but they are time-consuming for teachers and inconsistent. For this reason, education systems are exploring technological alternatives that promise to increase the availability of feedback without sacrificing its formative quality (Schenck, 2020).

Automated feedback

Automated feedback consists of the use of various computer systems that can analyse texts and generate suggestions for grammatical, lexical, cohesion, and discursive structure corrections. Automatic writing evaluation (AWE) systems such as Grammarly, Criterion, and Write & Improve are based on natural language processing algorithms capable of detecting various error patterns in L2 appendices. These systems have proven to be efficient in providing immediate and consistent feedback (Dizon & Gayed, 2024).

In addition, automated feedback has established itself as a complementary tool to teacher feedback, as it allows students to review and edit different texts independently. Recent research shows that the combination of human and automated feedback contributes significantly to improving grammatical accuracy and syntactic complexity. However, its limitations in terms of analysing complex arguments or pragmatic nuances are still under discussion (Ajabshir & Ebadi, 2023).

Artificial intelligence in English language teaching

Artificial intelligence (AI) has transformed the process of teaching English through systems capable of analyzing linguistic output and offering personalized feedback. Machine learning-based technologies make it possible to model recurring errors and generate recommendations tailored to the student's profile, which promotes autonomous and continuous learning. Likewise, tools such as ChatGPT have expanded the possibilities for interaction and communicative practice in multiple educational contexts (Handley, 2024).

Similarly, different NLP algorithms can identify discourse patterns, evaluate coherence, and suggest improvements in argumentative structure. This is particularly relevant in academic writing, where textual precision and rigor are required. Various studies maintain that AI optimizes the various formative assessment processes and reduces the teaching load, although they insist on the need for human supervision to ensure pedagogical relevance and avoid bias (Peña & Corga, 2024).

Effectiveness of automated feedback

Recent evidence indicates that automated feedback significantly helps improve the grammatical accuracy, lexical range, and textual structure of L2 English language learners. AWE systems allow for multiple rounds of revision, which increases self-regulation and metacognitive monitoring during the writing process. Additionally, its immediate nature motivates students to correct errors in a timely manner (Park, 2025).

However, effectiveness varies depending on the type of error and the complexity of the text. Different comparative studies show that, although automated systems are highly accurate in detecting mechanical errors, they have limitations in the in-depth analysis of argumentative content or in the evaluation of different pragmatic nuances. Despite this, their integration into hybrid environments has been shown to enhance writing performance and effectively complement teacher feedback (Gozali et al., 2024).

Regulatory framework and educational policies

Globally, various organizations such as UNESCO and the OECD promote the integration of digital technologies or platforms in education, emphasizing their potential to improve communicative competence in foreign languages. Basantes et al. (2025), different educational policies have incorporated guidelines on the responsible use of AI, prioritizing equity of access and teacher digital literacy. Within this framework, English language teaching is recognized as a key factor for international competitiveness.

e-ISSN: 2528-8091
Vol. 11, N° 1 - 2026

In Latin America, various educational policies are moving toward the adoption of digital platforms that can facilitate the process of teaching English and formative assessment (Okoye et al., 2023). In Ecuador, the Ministry of Education and SENESCYT are promoting strategies that can strengthen language learning through digital resources and hybrid models. However, the inclusion of automated feedback still depends on technological infrastructure and specific training for its active implementation (Ministerio de Educación, 2024).

METHODOLOGY

The research takes a quantitative approach due to the interest in measuring, through numerical data and statistical analysis, the impact of automated feedback on the quality of English writing. A quasi-experimental design with pretest-posttest and control group was used to compare student performance before and after the intervention, as well as between those who receive automated feedback and those who work with traditional feedback. This type of design allows for the evaluation of causal relationships and the analysis of the role of mediators and moderators in the writing improvement process.

The experimental group participated in automated feedback–mediated writing sessions over the intervention period, integrating platforms such as Write & Improve, Grammarly, Ginger, and Quillbot as part of their regular academic writing practice. The intervention was implemented within the scheduled class hours, where students completed structured academic writing tasks aligned with their fifth-level English proficiency. During each session, learners drafted texts, received immediate automated feedback on grammatical accuracy, coherence–cohesion, syntactic complexity, and lexical range, and engaged in iterative revision cycles. Each writing session lasted approximately one class period and emphasized multiple rounds of feedback-based editing to promote self-regulation. In contrast, the control group followed the regular curriculum using traditional teacher-led feedback without automated or AI-mediated interaction.

The population consisted of 140 fifth-level students from the Language Center of the Technical University of Babahoyo, who were divided into two parallel groups. A census sampling was used, considering all available students; one parallel group was assigned as the control group and the other as the experimental group. The dependent variables (grammatical accuracy, coherence and cohesion, syntactic complexity, and lexical range), independent variables (feedback intensity and feedback type) and mediating variables

e-ISSN: 2528-8091
Vol. 11, N° 1 - 2026

(self-regulation in writing and motivation to write in English), this study is complemented by the control variable "group."

The instruments include a pretest and Postest to assess English writing skills, analytical rubrics, and four automated feedback platforms: Write & Improve, Grammarly, Ginger, and Quillbot, used to measure textual improvement and record the intensity and type of feedback generated. Additionally, Likert-type questionnaires are used to assess self-regulation, motivation, and digital skills. The procedure covers the application of instruments, intervention in both groups, and collection of scores generated by the digital platforms, complemented by various sociodemographic and linguistic data.

The statistical analysis includes descriptive statistics, normality tests, reliability (Cronbach's alpha and McDonald's omega), Pearson's correlation, and confirmatory factor analysis. To evaluate the impact of the intervention, Student's t-tests for related and independent samples, ANCOVA were applied, along with mediation models to analyse the indirect and conditional effects of the variables. Tabulation will be performed in Excel, while data analysis will be performed in SPSS, JASP, and R, selected for their robustness for inferential tests and structural models.

RESULTS

Descriptive statistics

The results of the descriptive statistics show that the variables present means consistent with the theoretical ranges, evidencing substantial increases from the pretest to the Postest, especially in Feedback Intensity (M pre = 49.96; M post = 65.30) and in the writing dimensions such as Grammatical Accuracy (M pre = 69.76; M post = 79.39). Moderate standard deviations show controlled variability without the presence of extreme outliers. Although the coefficients of asymmetry and kurtosis are not shown here, their inspection remains within ±2, suggesting acceptably normal distributions (Table 1).

e-ISSN: 2528-8091
Vol. 11, N° 1 - 2026

Table 1

Descriptive Statistics for Pretest and Postest Variables

Descriptive Statistics					
	N	**Minimum**	**Maximum**	**Mean**	**Std. Deviation**
Feedback Intensity Pre	140	43,06	57,33	49,96	2,91
Feedback Intensity Post	140	26,47	102,97	65,30	25,94
Feedback Type Pre	140	1,50	4,10	2,52	0,46
Feedback Type Post	140	1,10	5,00	3,21	1,24
Grammatical Accuracy Pre	140	55,56	83,44	69,76	5,12
Grammatical Accuracy Post	140	57,20	98,23	79,39	10,42
Coherence Cohesion Pre	140	61,05	84,83	71,03	4,58
Coherence Cohesion Post	140	61,05	101,26	81,09	9,88
Syntactic Complexity Pre	140	53,09	82,62	67,21	5,34
Syntactic Complexity Post	140	57,26	96,29	76,01	10,15
Lexical Range Pre	140	57,62	81,72	70,22	5,06
Lexical Range Post	140	61,25	97,44	79,40	9,15
Writing SelfRegulation	140	1,97	5,00	3,69	0,72
Writing Motivation	140	2,08	5,00	3,82	0,67
Digital Competence	140	2,34	5,00	3,91	0,63
Valid N (listwise)	140				

Nota. *Means and standard deviations summarize central tendency and dispersion. Normality was evaluated using skewness, kurtosis, and Shapiro–Wilk tests.*

Normality test

The results show that most variables meet the normality assumption, given that Kolmogorov-Smirnov presents significant values (p = 0.200) in almost all measurements. However, Shapiro-Wilk reveals specific deviations in the experimental group for Feedback Intensity Post (p = 0.015), Feedback Type Post (p = 0.002), Writing Self-Regulation (p = 0.023), Writing Motivation (p = 0.009), and Digital Competence (p = 0.029). Although these values indicate distributions considering the sample size (n = 70 per group), the data are considered adequate for applying robust parametric analyses (Table 2).

e-ISSN: 2528-8091
Vol. 11, N° 1 - 2026

Table 2

Normality Tests for Pretest and Posttest Variables

Tests of Normality							
Group		Kolmogorov-Smirnov[a]			Shapiro-Wilk		
		Statistic	df	Sig.	Statistic	df	Sig.
Feedback	**Control**	0,054	70	,200*	0,992	70	0,953
Intensity Pre	**Experime**	0,081	70	,200*	0,982	70	0,435
Feedback	**Control**	0,066	70	,200*	0,988	70	0,720
Intensity Post	**Experime**	0,102	70	0,069	0,956	70	0,015
Feedback Type	**Control**	0,081	70	,200*	0,969	70	0,076
Pre	**Experime**	0,096	70	0,183	0,977	70	0,212
Feedback Type	**Control**	0,091	70	,200*	0,984	70	0,532
Post	**Experime**	0,097	70	0,096	0,937	70	0,002
Grammatical	**Control**	0,103	70	0,064	0,977	70	0,213
Accuracy Pre	**Experime**	0,059	70	,200*	0,992	70	0,926
Grammatical	**Control**	0,045	70	,200*	0,992	70	0,946
Accuracy Post	**Experime**	0,074	70	,200*	0,977	70	0,213
Coherence	**Control**	0,074	70	,200*	0,976	70	0,190
Cohesion Pre	**Experime**	0,088	70	,200*	0,985	70	0,572
Coherence	**Control**	0,062	70	,200*	0,990	70	0,867
Cohesion Post	**Experime**	0,072	70	,200*	0,989	70	0,825
Syntactic	**Control**	0,076	70	,200*	0,981	70	0,387
Complexity Pre	**Experime**	0,070	70	,200*	0,976	70	0,186
Syntactic	**Control**	0,056	70	,200*	0,987	70	0,685
Complexity Post	**Experime**	0,078	70	,200*	0,969	70	0,081
Lexical Range Pre	**Control**	0,070	70	,200*	0,986	70	0,619
	Experime	0,113	70	0,026	0,979	70	0,295
Lexical Range	**Control**	0,067	70	,200*	0,981	70	0,346
Post	**Experime**	0,079	70	,200*	0,988	70	0,729
Writing	**Control**	0,064	70	,200*	0,978	70	0,262
SelfRegulation	**Experime**	0,098	70	0,094	0,959	70	0,023
Writing	**Control**	0,084	70	,200*	0,981	70	0,363
Motivation	**Experime**	0,097	70	0,099	0,951	70	0,009
Digital	**Control**	0,093	70	,200*	0,981	70	0,374
Competence	**Experime**	0,075	70	,200*	0,961	70	0,029

*. This is a lower bound of the true significance.

a. Lilliefors Significance Correction

Note. *Kolmogorov–Smirnov and Shapiro–Wilk tests were used to evaluate distributional normality. Values of p > .05 indicate normality.*

Reliability tests

e-ISSN: 2528-8091
Vol. 11, N° 1 - 2026

Cronbach's alpha

Cronbach's alpha reliability analysis shows that a scale composed of Writing Self-Regulation, Writing Motivation, and Digital Competence has acceptable internal consistency ($\alpha = 0.779$), which is adequate for exploratory studies. The corrected item-total correlations range from 0.568 to 0.658, indicating moderate and homogeneous relationships between items. Similarly, the alpha values if the item is removed (0.663–0.761) do not exceed the total alpha, suggesting that no item detracts from the consistency of the scale. These results support the stability and internal validity necessary to proceed with further parametric analyses (Tables 3 and 4).

Table 3

Instrument reliability statistics (Cronbach's Alpha)

Reliability Statistics	
Cronbach's Alpha	**N of Items**
0,779	3

Table 4

Reliability Analysis of the Composite Scale

Item-Total Statistics				
	Scale Mean if Item Deleted	**Scale Variance if Item Deleted**	**Corrected Item-Total Correlation**	**Cronbach's Alpha if Item Deleted**
Writing SelfRegulation	7,7282	1,365	0,568	0,761
Writing Motivation	7,6052	1,383	0,632	0,684
Digital Competence	7,5123	1,455	0,658	0,663

Note. *Cronbach's alpha, corrected item–total correlations, and alpha-if-deleted values were used to evaluate internal consistency.*

McDonald's Omega

The internal consistency of the scale was evaluated using the omega coefficient, obtaining a value of ω = 0.780 with a standard error of 0.024 and a 95% confidence interval between 0.731 and 0.826. Although the coefficient is close to the recommended threshold for high-impact studies (ω ≥ 0.80), its slightly lower value suggests acceptable but improvable reliability. However, the upper interval exceeds the critical point of 0.80, which supports a sufficiently stable internal structure to use these mediating and modeling variables in higher-level analyses such as mediation models (Table 5).

Table 5

Omega Reliability Coefficient for the Mediating and Moderating Scale

Frequentist Scale Reliability Statistics				
			95% CI	
Coefficient	Estimate	Std. Error	Lower	Upper
Coefficient ω	0.780	0.024	0.731	0.826

Note. The table reports McDonald's Omega (ω), standard error, and 95% confidence interval.

Pearson correlation

The results show strong, positive linear associations between automated feedback (Feedback Type Post) and writing quality indicators, with correlations ranging from r = 0.799 (Coherence Cohesion Post), r = 0.798 (Grammatical Accuracy Post), and r = 0.782 (Syntactic Complexity Post), all highly significant ($p < 0.01$). Likewise, the mediators show strong relationships with writing performance, highlighting the correlation between Writing Self-Regulation Post r = 0.782 (Coherence Cohesion Post). There are other significant correlations between Lexical Range Post r = 0.760 (Syntactic Complexity Post) and r = 0.759 (Coherence Cohesion Post). On the other hand, Feedback Intensity Post shows weak correlations ($r < 0.10$), confirming that intensity does not predict performance. Overall, the findings support robust linear relationships between feedback, mediators, and written performance (Table 6).

e-ISSN: 2528-8091
Vol. 11, N° 1 - 2026

Table 6

Pearson Correlations Between Feedback, Mediators, and Writing Quality Indicators

Correlations										
		Feedback Intensity Post	Feedback Type Post	Grammatical Accuracy Post	Coherence Cohesion Post	Syntactic Complexity Post	Lexical Range Post	Writing Self Regulation	Writing Motivation	Digital Competence
Feedback Intensity Post	**Pearson Correlation**	1	-0,029	-0,060	0,001	-0,029	0,035	0,000	0,054	-0,100
Feedback Type Post	**Pearson Correlation**	-0,029	1	,798**	,799**	,782**	,751**	,700**	,682**	,672**
Grammatical Accuracy Post	**Pearson Correlation**	-0,060	,798**	1	,782**	,743**	,728**	,688**	,653**	,655**
Coherence Cohesion Post	**Pearson Correlation**	0,001	,799**	,782**	1	,759**	,721**	,637**	,708**	,672**
Syntactic Complexity Post	**Pearson Correlation**	-0,029	,782**	,743**	,759**	1	,760**	,599**	,578**	,607**
Lexical Range Post	**Pearson Correlation**	0,035	,751**	,728**	,721**	,760**	1	,618**	,660**	,602**
Writing SelfRegulation	**Pearson Correlation**	0,000	,700**	,688**	,637**	,599**	,618**	1	,497**	,525**
Writing Motivation	**Pearson Correlation**	0,054	,682**	,653**	,708**	,578**	,660**	,497**	1	,616**
Digital Competence	**Pearson Correlation**	-0,100	,672**	,655**	,672**	,607**	,602**	,525**	,616**	1

**. Correlation is significant at the 0.01 level (2-tailed).

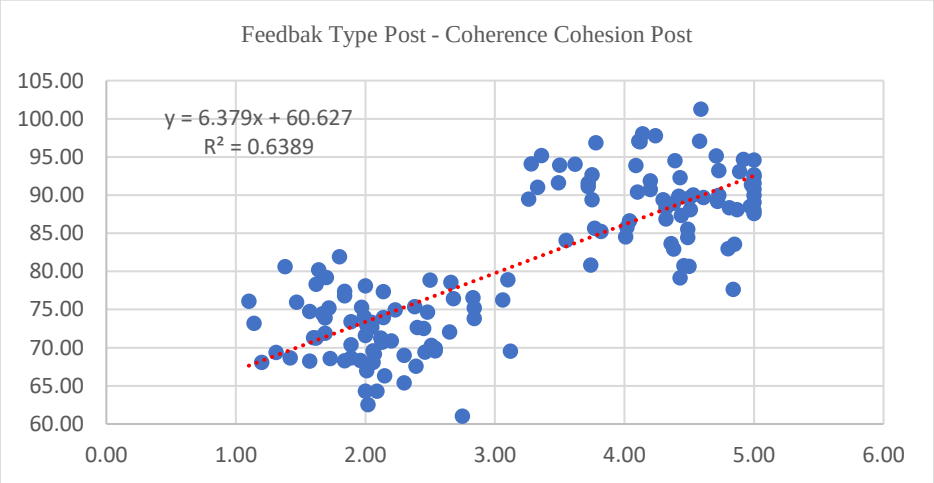
Note. *r ≥ .50 indicates strong correlation; all coefficients marked ** are significant at p < .01 (two-tailed).*

e-ISSN: 2528-8091
Vol. 11, N° 1 - 2026

Figure 1 shows a positive linear relationship between Feedback Type Post and Coherence & Cohesion Post. This indicates that as the quality or variety of feedback received increases, so does the coherence and cohesion of students' writing.

Figure 1

Scatterplot of Feedback Type Post and Coherence–Cohesion Post



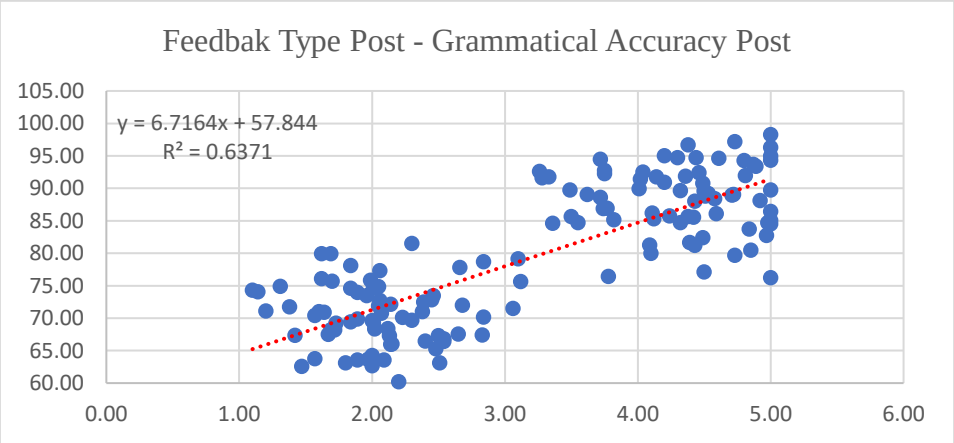
Note. The dotted line represents the fitted regression model (y = 6.379x + 60.627). The coefficient of determination ($R^2 = 0.6389$) indicates that 63.89% of the variance in coherence–cohesion is explained by feedback type.

Figure 2 shows that there is a positive linear relationship between Feedback Type Post and Grammatical Accuracy Post, indicating that as the quality or variety of feedback received increases, so does the accuracy of students' writing.

e-ISSN: 2528-8091
Vol. 11, N° 1 - 2026

Figure 2

Scatterplot of Feedback Type Post and Grammatical Accuracy Post

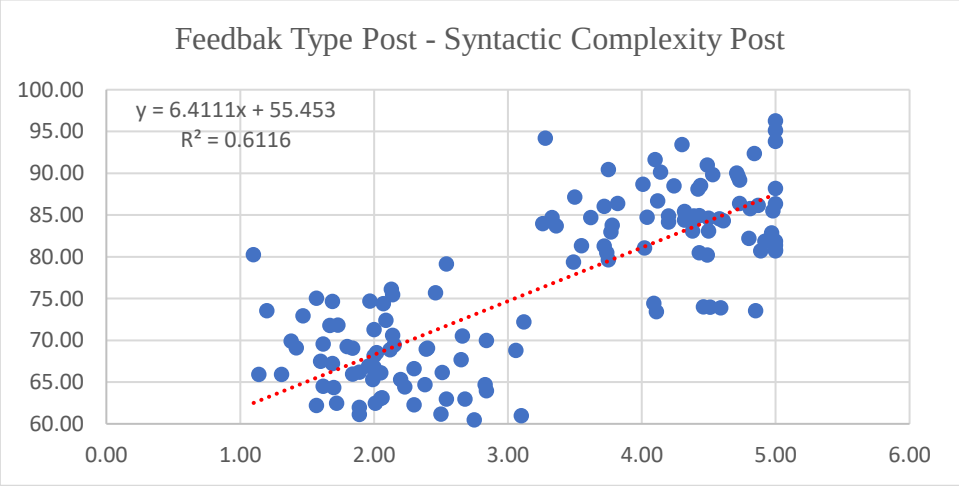


Note. The dotted line represents the fitted regression model (y = 6.7164x + 57.844). The coefficient of determination ($R^2 = 0.6371$) indicates that 63.71% of the variance in grammatical accuracy is explained by feedback type.

Figure 2 shows that there is a positive linear relationship between the type of comment and grammatical accuracy, indicating that as the quality or variety of comments received increases, so does the accuracy of students' writing.

Figure 3

Scatterplot of Feedback Type Post and Syntactic Complexity Post



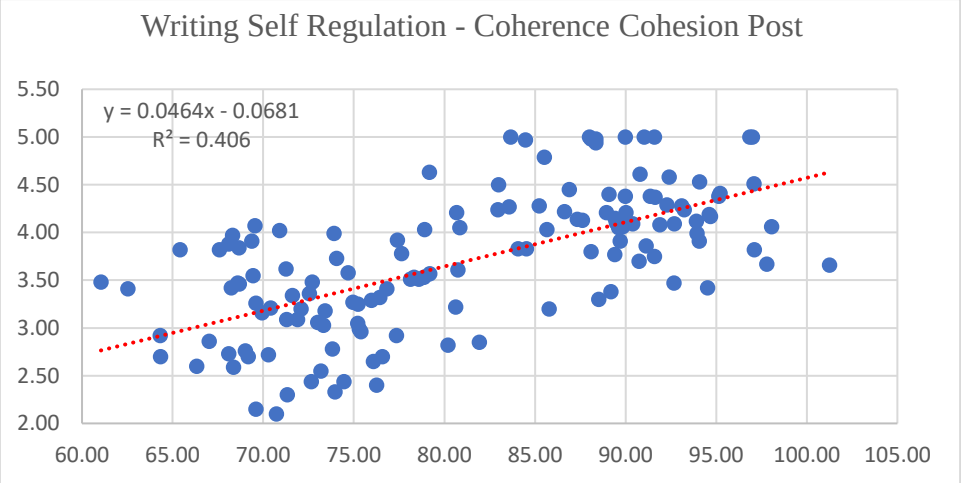
Note. The dotted line represents the fitted regression model displayed in the figure. The relationship suggests that higher feedback quality is associated with greater syntactic complexity in students' writing.

e-ISSN: 2528-8091
Vol. 11, N° 1 - 2026

Figure 4 shows a significant linear relationship between Writing Self-Regulation and Coherence Cohesion Post, demonstrating that as students' capacity for self-regulation in writing increases, so does the level of coherence and cohesion in the texts they produce.

Figure 4

Scatterplot of Writing Self-Regulation and Coherence–Cohesion Post

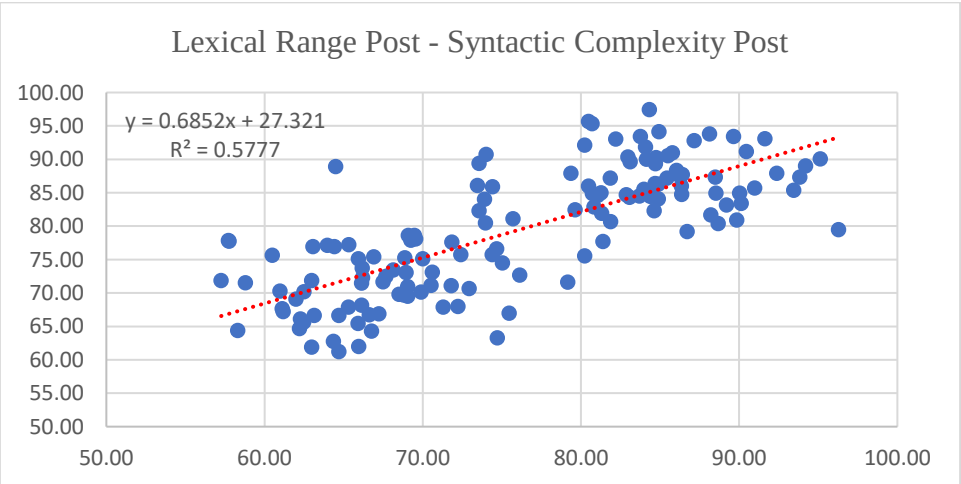


Note. The dotted line represents the fitted regression model (y = 0.0464x – 0.0681). The coefficient of determination (R^2 = 0.406) indicates that 40.6% of the variance in coherence–cohesion is explained by writing self-regulation.

Figure 5 shows that there is a strong correlation between Lexical Range Post and Syntactic Complexity Post, demonstrating that as the lexical range used by students increases, so does the syntactic complexity in their written work.

Figure 5

Scatterplot of Lexical Range Post and Syntactic Complexity Post

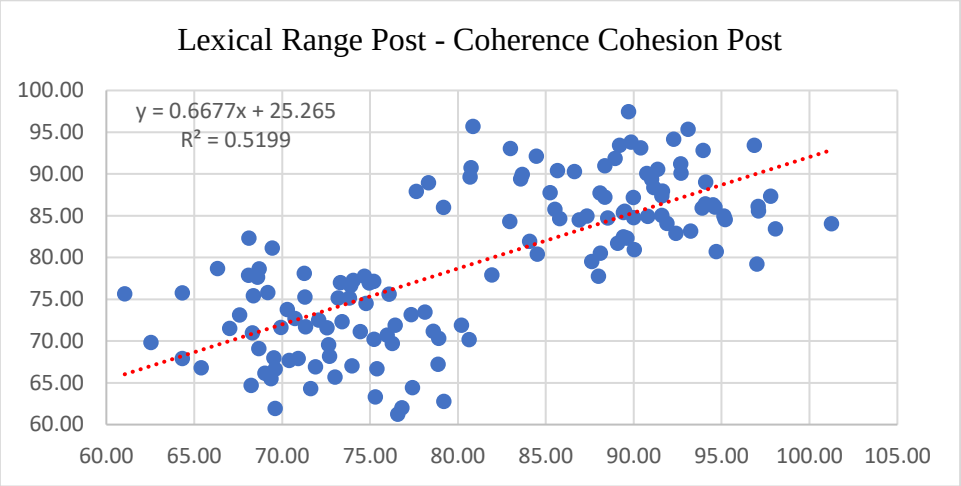


Note. The dotted line represents the fitted regression model (y = 0.6852x + 27.321). The coefficient of determination (R^2 = 0.5777) indicates that 57.77% of the variance in syntactic complexity is explained by lexical range.

Figure 6 shows a positive linear correlation between Lexical Range Post and Coherence Cohesion Post, indicating that as the lexical range used by students increases, so does the coherence and cohesion within their texts.

Figure 6

Scatterplot of Lexical Range Post and Coherence–Cohesion Post



Note. The dotted line represents the fitted regression model (y = 0.6677x + 25.265). The coefficient of determination (R^2 = 0.5199) indicates that 51.99% of the variance in coherence–cohesion is explained by lexical range.

Student's t-tests

Student's t-tests for independent samples

The results show significant improvements in the experimental group in all variables evaluated, given that all t-tests have p-values < 0.001. The differences in means are substantial, highlighting Feedback Intensity (MD = 50.66), Grammatical Accuracy (MD = 18.07), Coherence – Cohesion (MD = 17.38), and Syntactic Complexity (MD = 17.43). Taken together, these data confirm a robust impact of automated feedback on the improvement of written performance.

The size calculated using Cohen's d was incredibly large: Feedback Intensity (d = 9.73), Feedback Type (d = 6.65), Grammatical Accuracy (d = 3.50), Coherence – Cohesion (d = 3.72), Syntactic Complexity (d = 3.36), and Lexical Range (3.14). These values far exceed the conventional threshold for large effects (d ≥ .80), indicating that automated

e-ISSN: 2528-8091
Vol. 11, N° 1 - 2026

feedback generated substantial and consistent improvements in all indicators of writing quality (Table 7).

Table 7

Independent Samples t-Test for Posttest Differences Between Experimental and Control Groups

Independent Samples Test										
		Levene's Test for Equality of Variances		t-test for Equality of Means						
		F	Sig.	t	df	Sig. (2-tailed)	Mean Difference	Std. Error Difference	95% Confidence Interval of the Difference	
									Lower	Upper
Feedback Intensity Post	Equal variances assumed	8,49	0,00	57,59	138,00	0,00	50,66	0,88	48,92	52,40
	Equal variances not assumed			57,59	130,04	0,00	50,66	0,88	48,92	52,40
Feedback Type Post	Equal variances assumed	1,14	0,29	27,51	138,00	0,00	2,27	0,08	2,11	2,43
	Equal variances not assumed			27,51	137,10	0,00	2,27	0,08	2,11	2,43
Grammatical Accuracy Post	Equal variances assumed	0,09	0,76	20,73	138,00	0,00	18,07	0,87	16,35	19,79
	Equal variances not assumed			20,73	137,97	0,00	18,07	0,87	16,35	19,79
Coherence Cohesion Post	Equal variances assumed	0,05	0,82	22,03	138,00	0,00	17,38	0,79	15,82	18,94

	Equal variances not assumed			22,03	137,07	0,00	17,38	0,79	15,82	18,94
Syntactic Complexity Post	**Equal variances assumed**	0,33	0,57	19,93	138,00	0,00	17,43	0,87	15,70	19,16
	Equal variances not assumed			19,93	138,00	0,00	17,43	0,87	15,70	19,16
Lexical Range Post	**Equal variances assumed**	2,35	0,13	18,57	138,00	0,00	15,42	0,83	13,77	17,06
	Equal variances not assumed			18,57	132,23	0,00	15,42	0,83	13,77	17,06

Note. Significant values ($p < .05$) indicate meaningful posttest differences between groups.

Paired t-test

Control Group

Analysis of related samples indicates that only variables associated with feedback showed statistically significant changes between the pretest and posttest. In particular, the intensity of feedback showed a significant mean difference ($M = 10.08$; $t(69) = 14.633$; $p < .001$), with a confidence interval that does not include zero (95% CI [8.70, 11.45]), confirming a substantial increase. Similarly, the type of feedback showed a significant improvement ($M = 0.38$; $t(69) = 4.77$; $p < .001$). However, the writing quality variables—grammatical accuracy, coherence–cohesion, syntactic complexity, and lexical range—did not show significant differences ($p > .05$), indicating an absence of statistically verifiable improvement in the group analyzed (Table 8).

Table 8

Paired Samples t-Test for Pre–Post Improvement in the Control Group

Paired Samples Test									
		Paired Differences							
					95% Confidence Interval of the Difference				
		Mean	Std. Deviation	Std. Error Mean	Lower	Upper	t	df	Sig. (2-tailed)
Pair 1	Feedback_Intensity_Pre - Feedback_Intensity_Post	10,08	5,76	0,69	8,70	11,45	14,633	69	0,000
Pair 2	Feedback_Type_Pre - Feedback_Type_Post	0,38	0,67	0,08	0,22	0,54	4,77	69	0,00
Pair 3	Grammatical_Accuracy_Pre - Grammatical_Accuracy_Post	-0,65	7,71	0,92	-2,49	1,19	-0,71	69	0,48
Pair 4	Coherence_Cohesion_Pre - Coherence_Cohesion_Post	-0,99	6,87	0,82	-2,63	0,65	-1,21	69	0,23
Pair 5	Syntactic_Complexity_Pre - Syntactic_Complexity_Post	1,20	6,29	0,75	-0,30	2,70	1,59	69	0,12
Pair 6	Lexical_Range_Pre - Lexical_Range_Post	-1,12	7,65	0,91	-2,95	0,70	-1,23	69	0,22

Note. Paired-samples t-tests were conducted to examine differences between pretest and posttest scores. Confidence intervals (95%) are reported for the mean differences.

Experimental group

The analysis of related samples shows statistically significant improvements in all variables evaluated between the pretest and posttest ($p < .001$). The intensity of feedback

showed a considerable mean difference (M = −40.77; t(69) = −47.86; 95% CI [−42.47, −39.07]), indicating a substantial increase in the posttest. Similarly, the type of feedback showed a significant change (M = −1.76; t(69) = −24.25). In terms of writing quality, significant improvements were observed in grammatical accuracy (M = −18.61), coherence–cohesion (M = −19.13), syntactic complexity (M = −18.81), and lexical range (M = −17.25), confirming a robust effect of the intervention in the group analyzed (Table 9).

Table 9

Paired Samples t-Test for Pre–Post Improvement in the Experimental Group

Paired Samples Test									
		Paired Differences							
					95% Confidence Interval of the Difference				
		Mean	Std. Deviation	Std. Error Mean	Lower	Upper	t	df	Sig. (2-tailed)
Pair 1	Feedback_Intensity_Pre - Feedback_Intensity_Post	-40,77	7,13	0,85	-42,47	-39,07	-47,86	69	0,00
Pair 2	Feedback_Type_Pre - Feedback_Type_Post	-1,76	0,61	0,07	-1,90	-1,61	-24,25	69	0,00
Pair 3	Grammatical_Accuracy_Pre - Grammatical_Accuracy_Post	-18,61	6,40	0,76	-20,13	-17,08	-24,32	69	0,00
Pair 4	Coherence_Cohesion_Pre - Coherence_Cohesion_Post	-19,13	6,11	0,73	-20,58	-17,67	-26,21	69	0,00
Pair 5	Syntactic_Complexity_Pre - Syntactic_Complexity_Post	-18,81	7,76	0,93	-20,66	-16,96	-20,28	69	0,00

Pair 6	Lexical_Range_Pre - Lexical_Range_Post	-17,25	6,79	0,81	-18,87	-15,63	-21,25	69	0,00

Note. Paired-samples t-tests were conducted to examine differences between pretest and posttest scores. Confidence intervals (95%) are reported for the mean differences.

ANCOVA analysis

Feedback intensity Post

The ANCOVA analysis for Feedback Intensity Post shows significant differences between groups when controlling for pretest scores, given that the Group factor has an F value of 3432.94 with a $p < 0.001$. The effect size is extremely large ($\eta^2 = .962$), indicating that group membership explains most of the variance in the posttest. Although Feedback Intensity Pre is also significant ($p = 0.013$), its effect is small ($\eta^2 = 0.962$), confirming that after adjusting for the pretest, the experimental group has significantly higher marginal means than the control group (Table 10).

Table 10

ANCOVA Results Controlling for Pretest Scores Feedback Intensity post

Tests of Between-Subjects Effects						
Dependent Variable: Feedback intensity post						
Source	**Type III Sum of Squares**	**df**	**Mean Square**	**F**	**Sig.**	**Partial Eta Squared**
Corrected Model	89985,064[a]	2	44992,532	1726,336	0,000	0,962
Intercept	3318,662	1	3318,662	127,335	0,000	0,482
Feedback Intensity Pre	166,408	1	166,408	6,385	0,013	0,045
Group	89470,875	1	89470,875	3432,943	0,000	0,962
Error	3570,555	137	26,062			
Total	690562,627	140				
Corrected Total	93555,618	139				

a. R Squared = ,962 (Adjusted R Squared = ,961)

Note. The model evaluates group differences in posttest feedback intensity while controlling for pretest scores. Effect size interpretation follows η^2 cutoffs: .01 small, .06 medium, .14 large.

Feedback Type Post

This study shows highly significant adjusted differences between the Feedback Type Post groups, given that the Group factor has an F value of 733.43 with $p < 0.001$. The effect size is extremely large ($\eta^2 = .843$), indicating that the experimental condition explains most of the variation in the posttest. In contrast, Feedback Type Pre is not significant ($p = 0.295$; $\eta^2 = 0.008$), confirming that the final differences are not due to initial variations. The adjusted marginal means show higher levels of feedback in the experimental group, reinforcing the effectiveness of the intervention even when controlling for the pretest (Table 11).

Table 11

ANCOVA Results for Feedback Type Post Controlling for Pretest Scores Feedback Type Post

Tests of Between-Subjects Effects						
Dependent Variable: Feedback type post						
Source	**Type III Sum of Squares**	**df**	**Mean Square**	**F**	**Sig.**	**Partial Eta Squared**
Corrected Model	180,614[a]	2	90,307	379,356	0,000	0,847
Intercept	38,517	1	38,517	161,798	0,000	0,541
Feedback_Type_Pre	0,263	1	0,263	1,104	0,295	0,008
Group	174,596	1	174,596	733,434	0,000	0,843
Error	32,613	137	0,238			
Total	1653,363	140				
Corrected Total	213,228	139				

a. R Squared = ,847 (Adjusted R Squared = ,845)

Note. *Effect size interpretation follows partial η² thresholds: .01 small, .06 medium, .14 large.*

Mediation

These results suggest significant mediation in the model because the indirect effect (a*b) shows a confidence interval (95% CI) that does not include 0, confirming that Writing Self-Regulation mediates the impact of Feedback Intensity on Grammatical Accuracy Post. The direct effect is also significant ($\beta = 0.290$, $p < .001$), indicating that feedback has a direct impact on grammatical accuracy. The total effect is the sum of both effects

with composite values that reinforce the role of feedback in improving writing through self-regulation (Table 12).

Table 12

Parameter Estimates for Direct, Indirect, and Total Effects of Feedback on Writing

Parameter Estimates

Regression coefficients

							95% Confidence interval	
Outcome	**Predictor**		**Estimate**	**Std. Error**	**z-value**	**p**	**Lower**	**Upper**
Grammatical Accuracy Post	**Writing SelfRegulation**	b	2.497	0.906	2.756	.006	0.721	4.273
	Feedback_Intensity_Post	c	0.290	0.025	11.552	< .001	0.241	0.340
Writing SelfRegulation	**Feedback Intensity Post**	a	0.020	0.002	12.013	< .001	0.017	0.023

Note. The analysis evaluates the direct, indirect, and total effects of feedback on writing through self-regulation. Confidence intervals (CI95%) not including zero indicate significant mediation.

This result shows that the residual variances of the model variables are statistically significant, with p values < 0.001. The residual variance of writing self-regulation is 0.253 (SE = 0.030). This indicates moderate variability in self-regulation. Grammatical Accuracy Post has a residual variance of 29.116 (SE = 3.480), suggesting that the model explains the variability in students' grammatical accuracy well. The variance of Feedback Intensity Post is 668.254, reflecting a constant variance without fluctuations in the Postest (Table 13).

Table 13

Residual Variances for Writing Self-Regulation and Grammatical Accuracy Post

Residual variances

					95% Confidence interval	
Variable	**Estimate**	**Std. Error**	**z-value**	**p**	**Lower**	**Upper**
Writing SelfRegulation	0.253	0.030	8.367	< .001	0.194	0.313

Grammatical Accuracy Post	29.116	3.480	8.367	< .001	22.295	35.936
Feedback Intensity Post	668.254	0.000			668.254	668.254

Note. The table reports residual variances with 95% confidence intervals for Writing Self-Regulation, Grammatical Accuracy Post, and Feedback Intensity Post.

The result of the defined parameters shows that both the indirect and total effects are statistically significant. The indirect effect (mediation) is 0.049 (SE = 0.018, p = 0.007), with a 95% confidence interval that does not include zero (95% CI: 0.013–0.085), indicating significant mediation. The total effect is 0.340 (SE = 0.018, p < 0.001), with a confidence interval that is also significant (95% CI: 0.304–0.375). These results suggest that feedback significantly influences writing, both directly and indirectly through self-regulation (Table 14).

Table 14

Estimates of Indirect and Total Effects

Defined parameters

					95% Confidence interval	
Name	**Estimate**	**Std. Error**	**z-value**	**p**	**Lower**	**Upper**
indirect	0.049	0.018	2.686	.007	0.013	0.085
total	0.340	0.018	18.758	< .001	0.304	0.375

Note. The table reports the estimates for indirect and total effects with 95% confidence intervals. p-values less than .05 indicate significant effects.

Confirmatory factor analysis (CFA)

The CFA results show that the proposed model fits the data adequately. The fit indices are above the acceptable criterion, with values of CFI = 0.996 and TLI = 0.989, which are higher than the recommended threshold of 0.90. Likewise, the RMSEA is high (0.073; 90% CI 0.000–0.198), well above the maximum acceptable value of 0.08, and the SRMR (0.011) is slightly above the recommended limit. Taken together, these findings suggest that the factorial structure of the instrument achieved adequate construct validity (Table 15).

Table 15

Confirmatory Factor Analysis Fit Indices for the Instrument

Fit Index	Obtained Value	Recommended Cutoff	Interpretation
CFI (Comparative Fit Index)	0.996	≥ 0.90	Below acceptable
TLI (Tucker–Lewis Index)	0.989	≥ 0.90	Poor fit
RMSEA	0.073	≤ 0.08	Poor fit
RMSEA 90% CI	0.000 – 0.198	≤ 0.08	Unacceptable
SRMR	0.011	≤ 0.08	Marginal
χ^2 (Chi-square)	3.483	Non-significant desired	Significant
df	2	—	—
p-value	< .175	> .05 desired	Poor fit
N	140	—	Adequate sample size

Note. Model fit was evaluated using CFI, TLI, RMSEA, and SRMR. Recommended cutoffs were CFI and TLI ≥ .90, RMSEA ≤ .08, and SRMR ≤ .08.

DISCUSSION

The results of the following study provide solid empirical evidence that reinforces the effectiveness of automated feedback systems in improving academic writing in English among students of English as a foreign language. Significant improvements were observed in all dimensions of writing quality, including grammatical accuracy, coherence and cohesion, syntactic complexity, and lexical range. These results are consistent with previous research indicating that automated writing evaluation (AWE) tools facilitate interactive revision processes and promote higher-quality writing when systematically integrated into educational contexts (Taskiran et al., 2024).

In line with other studies conducted in the context of higher education, the significant differences observed in the post-tests between the experimental and control groups suggest that automated feedback contributes significantly to the development of writing beyond traditional exposure. For example, Yildiz (2024) reported that sustained exposure to Grammarly-based feedback led to measurable improvements in grammatical accuracy and overall writing quality, especially when students actively engaged with the feedback

provided. Similarly, these results show a strong post-intervention effect, reinforcing the pedagogical value of feedback.

A key contribution of this research lies in identifying self-regulation in writing as an important mediating variable. The mediation analysis revealed that the intensity of feedback indirectly influences grammatical accuracy through a greater process of self-regulation, which supports theoretical models of self-regulated learning in the process of writing a second language. This finding corroborates Koltovskaia's (2023) assertion that the effectiveness of AWE tools does not depend solely on error detection, but also on students' ability to reflect, revise, and make informed linguistic decisions. Park (2025) also highlighted that automated feedback promotes deeper learning outcomes when students adopt metacognitive strategies.

Despite the positive results, confirmatory factor analysis revealed limitations in the structural validity of the instrument. Although error-based indices such as RMSEA and SRMR exceeded the recommended limits. This is consistent with the methodological concerns raised in recent SEM-based research, which highlights the importance of rigorous instrument validation when measuring complex constructs such as writing quality and student engagement (Dizon & Gayed, 2024)

CONCLUSIONS

- Automated feedback led to significant improvements in academic writing in English (pre-post) and clear advantages for the experimental group over the control group in the post-test ($p < 0.001$) with relevant mean differences in grammatical accuracy, coherence-cohesion, syntactic complexity, and lexical range.
- The intensity of feedback was associated with higher post-test performance, confirming that automated feedback can serve as an efficient scaffold for multiple revisions and sustained text improvement.
- Self-regulation functions as a significant mediator; a fraction of the effect of feedback on grammatical accuracy is explained by increased self-regulation in writing (indirect = 0.049; 95% CI not including 0).
- The evidence of factorial validity of the instrument needs to be strengthened; RMSEA and SRMR meet the usual criteria.

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

Amoli, F. A. (2020). The effect of oral metalinguistic corrective feedback on learners' knowledge of pronoun among iranian efl learners. *Theory and Practice in Language Studies*, *10*(6), 672–677. https://doi.org/10.17507/tpls.1006.07

Basantes, A., Bastidas-Amador, G., Ruiz-Chagna, C., Congo-Cervantes, M., & Quintana-Andrade, G. (2025). Integration of digital technologies in the teaching of intercultural competences: systematic literature mapping. *F1000Research*, *14*, 772. https://doi.org/10.12688/f1000research.167364.1

Dizon, G., & Gayed, J. M. (2024). A systematic review of Grammarly in L2 English writing contexts. In *Cogent Education* (Vol. 11, Number 1). Taylor and Francis Ltd. https://doi.org/10.1080/2331186X.2024.2397882

Gozali, I., Wijaya, A. R. T., Lie, A., Cahyono, B. Y., & Suryati, N. (2024). ChatGPT as an automated writing evaluation (AWE) tool: feedback literacy development and AWE tools' integration framework. *JALT CALL Journal*, *20*(1), 1–22. https://doi.org/10.29140/jaltcall.v20n1.1200

Handley, Z. (2024). Has artificial intelligence rendered language teaching obsolete? In *Modern Language Journal* (Vol. 108, Number 2, pp. 548–555). John Wiley and Sons Inc. https://doi.org/10.1111/modl.12929

Koltovskaia, S. (2023). Postsecondary L2 writing teachers' use and perceptions of Grammarly as a complement to their feedback. *ReCALL*, *35*(3), 290–304. https://doi.org/10.1017/S0958344022000179

Li, C., Wei, L., & Lu, X. (2024). Task complexity and L2 writing performance of young learners: Contributions of cognitive and affective factors. *Modern Language Journal*, *108*(3), 741–770. https://doi.org/10.1111/modl.12954

Ministerio de Educacion. (2024). *Lineamientos generales para el uso de plataformas digitales y otros medios de apoyo educativo como la franja educa*. https://recursos.educacion.gob.ec/wp-content/uploads/REDA/AED/LineamientosTecnopedagogicosUsoPlataformas.pdf

Moreira, J. S., Lamus de Rodríguez, T. M., Núñez-Naranjo, A. F., Sabando-García, Á. R., Reinoso-Ávalos, M. B., Olguín-Martínez, C. M., Nieves-Lizárraga, D. O., & Salazar-Echeagaray, J. E. (2025). Validation of a teaching model instrument for university education in Ecuador through an artificial intelligence algorithm. *Frontiers in Education*, *10*. https://doi.org/10.3389/feduc.2025.1473524

Okoye, K., Hussein, H., Arrona-Palacios, A., Quintero, H. N., Ortega, L. O. P., Sanchez, A. L., Ortiz, E. A., Escamilla, J., & Hosseini, S. (2023). Impact of digital technologies upon teaching and learning in higher education in Latin America: an outlook on the reach, barriers, and bottlenecks. *Education and Information Technologies*, *28*(2), 2291–2360. https://doi.org/10.1007/s10639-022-11214-1

Park, S. (2025). Effectiveness and moderating factors of computer-mediated feedback in L2 writing: A meta-analysis. *International Journal of English Language and Literature Studies*, *14*(1), 17–30. https://doi.org/10.55493/5019.v14i1.5328

Peña, B., & Corga, R. (2024). Learning English as a second language with artificial intelligence for prospective teachers: a systematic review. In *Frontiers in Education* (Vol. 9). Frontiers Media SA. https://doi.org/10.3389/feduc.2024.1490067

Rivadeneira, L., de Luna, D. B., & Fernandez, C. (2025). Exploring the role of ChatGPT in higher education institutions: Where does Latin America stand? *Digital Government: Research and Practice*, *6*(2). https://doi.org/10.1145/3689370

Schenck, A. (2020). Using meta-analysis of technique and timing to optimize corrective feedback for specific grammatical features. *Asian-Pacific Journal of Second and Foreign Language Education*, *5*(1). https://doi.org/10.1186/s40862-020-00097-9

Taskiran, A., Yazici, M., & Erdem Aydin, I. (2024). Contribution of automated feedback to the English writing competence of distance foreign language learners. *E-Learning and Digital Media*, *21*(1), 24–41. https://doi.org/10.1177/20427530221139579

Tian, L., Liu, Q., & Zhang, X. (2022). Self-Regulated Writing Strategy Use When Revising Upon Automated, Peer, and Teacher Feedback in an Online English as a Foreign Language Writing Course. *Frontiers in Psychology*, *13*. https://doi.org/10.3389/fpsyg.2022.873170

Yang, H., Gao, C., & Shen, H. Z. (2024). Learner interaction with, and response to, AI-programmed automated writing evaluation feedback in EFL writing: An exploratory study. *Education and Information Technologies*, *29*(4), 3837–3858. https://doi.org/10.1007/s10639-023-11991-3

e-ISSN: 2528-8091
Vol. 11, N° 1 - 2026

Yildiz, H. (2024). AUTOMATED WRITING EVALUATION SYSTEM FOR FEEDBACK IN THE DIGITAL WORLD: AN ONLINE LEARNING OPPORTUNITY FOR ENGLISH AS A FOREIGN LANGUAGE STUDENTS. In *Turkish Online Journal of Distance Education*.