Available online at https://www.ejal.info
http://dx.doi.org/10.32601/ejal.11311

Eurasian Journal of Applied Linguistics, 11(3) (2025) 114-127

EJAL
Eurasian Journal of Applied Linguistics

Integrating AI-Based Feedback into EFL Writing Instruction: A Case Study from a Vietnamese University

Le Thi Minh Thao[a*], Hoang Tuyet Minh[b], Dinh Tuan Long[c], Nguyen Thi Ngoc Anh[d]

[a] *Department for Scientific Research and External Relations, Hanoi Open University, Vietnam. Email: thaoltm@hou.edu.vn*

[b] *International School, Vietnam National University, Vietnam. Email: tuyetminh@vnu.edu.vn*

[c] *Faculty of Information Technology, Hanoi Open University, Vietnam Hanoi Open University, Vietnam Email: dinhtuanlong@hou.edu.vn*

[d] *Technology for Industries, Vietnam. Email: ntnanh@uneti.edu.vn*

Received: 20 June 2025 | Received: in Revised Form 05 August 2025 | Accepted 25 August 2025

APA Citation:
Thao, L. T. M, Minh, H. T., Long, D. T., Anh, N. T. N. (2025). Integrating AI-Based Feedback into EFL Writing Instruction: A Case Study from a Vietnamese University *Eurasian Journal of Applied Linguistics, 11*(3), 114-127.
Doi: http://dx.doi.org/10.32601/ejal.11311

Abstract
There exists inadequacy of traditional feedback mechanisms in EFL writing instruction within Vietnamese higher education contexts. This research aims to investigate the effectiveness and feasibility of implementing AI-based feedback systems in enhancing Vietnamese university students' EFL writing skills. Mixed-methods sequential explanatory design was employed over a 16-week period. Random sampling was used to assign 384 Hanoi Open University second and third-year undergraduate students to experimental (AI feedback, n = 192) and control (traditional teacher feedback, n = 192) groups. Complexity, Accuracy, and Fluency (CAF) measures were used to evaluate argumentative writing tests administered before and after intervention, supplemented by semi-structured interviews and technology acceptance questionnaires. The experimental group demonstrated significantly larger improvements in all CAF dimensions, with significant gains in fluency (d = 0.51-0.61), moderate effects for complexity (d = 0.45-0.49), and large effect sizes for accuracy (Cohen's d = 0.66-0.71). Students showed a 5.6-point decrease in errors per 100 words and an 18.2% increase in error-free T-units specifically. Although there were limitations in terms of cultural sensitivity (65% of participants) and enduring needs for human interaction (87.5% of participants), student acceptance was high (M = 3.97/5.0). Results demonstrate that AI-based feedback systems significantly enhance EFL writing when integrated with traditional teaching methods, providing scalable solutions for large university settings while requiring targeted implementation strategies and comprehensive teacher training in hybrid pedagogies.

Keywords: Artificial Intelligence, EFL Writing, Automated Feedback, Higher Education, Language Learning Technology, CAF Framework

* Corresponding Author.
Email: thaoltm@hou.edu.vn
DOI: http://dx.doi.org/10.32601/ejal.11311

Introduction

Higher education institutions experience a transformative change in their pedagogical practices because of artificial intelligence technology integration across various academic disciplines. According to Akinwalere and Ivanov (2022) higher education institutions face both new opportunities and challenges through AI which helps improve personalized learning and address scalability issues. The UNESCO report from Pedro et al. (2019) shows how AI technology can support various learning needs while ensuring that everyone has access to high-quality education. Since traditional teaching methods are unable to give a personalized student feedback, the emerging developments bring about significant changes for language education. The EFL instruction in Vietnamese higher education institutions faces distinctive obstacles which become most apparent in large educational settings. The development of English writing skills represents one of the most demanding cognitive abilities which needs extensive personalized feedback and individualized attention.

This study specifically utilized ChatGPT-4 as the primary AI tool for generating automated feedback on student writing, integrated with Grammarly's automated writing evaluation system for comprehensive error detection and correction suggestions. These AI tools were integrated into the pedagogical framework through a hybrid approach where students received immediate automated feedback on grammatical accuracy, vocabulary usage, and structural organization, while maintaining traditional teacher guidance for creativity, cultural context, and complex argumentative development. The EFL education system in Vietnam operates with limited resources because many universities deal with high student-to-instructor ratios which make complete writing instruction challenging. The large student enrollment at Hanoi Open University (HOU) with 30,000 students across 18 undergraduate programs makes it difficult to maintain traditional teacher-centered feedback methods at an institutional scale.

The central problem addressed in this study is the inadequacy of traditional feedback mechanisms in EFL writing instruction within Vietnamese higher education contexts, particularly in large-scale institutional settings. Multiple studies have shown that traditional feedback approaches used in EFL writing instruction have significant limitations. Ferris (2014) shows that teachers face three main challenges when providing complete feedback because of time constraints and big class sizes and the difficulty of providing consistent responses. Lee (2017) shows that classroom writing assessment practices are limited by institutional factors and resource limitations and that feedback interventions fail to achieve their intended outcomes when delivered inconsistently. The challenges at HOU become more severe because of its large size and diverse student body, creating a critical gap between pedagogical needs and available resources that demand immediate attention and innovative solutions.

The selection of AI-based feedback systems for EFL writing instruction is motivated by several compelling factors. First, the urgent need for scalable educational solutions in Vietnam's rapidly expanding higher education sector necessitates exploration of technology-enhanced pedagogical approaches. Second, the proven limitations of traditional feedback methods in large classroom settings create an imperative for alternative approaches that can maintain quality while serving increased student populations. Third, the global advancement in AI technologies presents unprecedented opportunities for language education that remain underexplored in Vietnamese contexts. Finally, the alignment between AI capabilities in natural language processing and the specific needs of writing instruction makes this technological application particularly promising for addressing current educational challenges.

This research aims to investigate the effectiveness and feasibility of implementing AI-based feedback systems in EFL writing instruction within Vietnamese higher education contexts. Through its assessment of useful AI technology applications in extensive EFL writing instruction, the research significantly advances Computer-Assisted Language Learning (CALL). By demonstrating how new technologies can enhance educational outcomes rather than replace human teaching methods, the research explores fundamental questions regarding the scalability of technology-enhanced language learning. The study supports Vietnam's educational digital transformation strategy by offering evidence-based suggestions for instructional technologists and university administrators seeking scalable writing instruction solutions that directly address the identified problems of resource limitations and inadequate personalized feedback provision.

To address these challenges and bridge the gap between traditional limitations and modern educational needs, three primary questions are examined by the study: (1) In comparison to conventional techniques, how much does AI-based feedback enhance EFL students' writing performance in terms of accuracy, complexity, and fluency? (2) What are students' attitudes and perceptions toward AI-based writing feedback systems? (3) What factors influence the effectiveness of AI-based feedback interventions? The investigation pursues three interconnected objectives that directly respond to the identified problems: evaluating AI feedback effectiveness on CAF measures to address the limitation of inconsistent traditional feedback, investigating student perceptions and acceptance to ensure pedagogical viability in Vietnamese contexts, and identifying factors that influence intervention effectiveness to provide guidance for optimizing AI-based feedback systems across diverse student populations and institutional contexts thereby offering scalable solutions to resource constraints.

Literature Review and Theoretical Framework

Second Language Writing Theory

The theoretical foundation for understanding the evolution of second language writing is comprised of several interconnected frameworks that provide insight into the complex mechanisms behind L2 composition. A comprehensive examination of writing in a second language, emphasizing the intricacy of writing proficiency, which encompasses linguistic, cognitive, and social elements was provided by Hyland (2019). This research showed that in addition to teaching surface-level linguistic features, effective L2 writing instruction must address higher-order cognitive processes like planning, organizing, and revising textual content. Recent empirical validation of these theoretical principles comes from Song and Song (2023) mixed-methods study involving 50 Chinese EFL students, which demonstrated significant improvements in writing organization, coherence, grammar, and vocabulary through AI-assisted instruction. Their quantitative analysis revealed substantial enhancements in participants' academic writing performance ($d = 0.76$, large effect size) compared to traditional instruction methods, providing concrete evidence for the multi-dimensional nature of L2 writing development proposed by Hyland.

The complexity, accuracy, and fluency (CAF) framework is one of the core theoretical ideas in second language writing research. Pallotti (2020) explains the theoretical underpinnings and practical applications of CAF measures in second language acquisition and language testing contexts. The framework provides a systematic approach to evaluating writing development in three key areas: complexity refers to the range and complexity of linguistic structures employed; fluency refers to the ease and speed of language processing and production; and accuracy includes the degree of error-free language production. This tripartite framework allows researchers and practitioners to look at writing improvement and instructional effectiveness from a broad perspective. The practical application of CAF measures in AI-enhanced writing contexts has been empirically validated by recent studies. Sari and Han (2024) conducted a comprehensive investigation examining the impact of automated writing evaluation on 56 EFL learners' writing performance using CAF measures. Their findings revealed significant improvements in grammatical accuracy ($F(1,47) = 9.86$, $p = 0.01$) and structural complexity following AI-based feedback intervention. Similarly, Shen et al. (2023) applied CAF measures in their analysis of 120 EFL learners' writing development under computer-generated feedback instruction, finding no significant differences between treatment and control groups on complexity and accuracy measures, highlighting the nuanced nature of AI feedback effectiveness.

Theory of Feedback in L2 Writing

The theoretical landscape of feedback in second language writing has expanded dramatically following decades of empirical research and theoretical development. By providing a comprehensive examination of written corrective feedback for L2 development, Bitchener and Storch (2016) establish the theoretical framework for connecting feedback delivery to broader theories of second language acquisition. According to their research, feedback must be timely, focused, and appropriate for the students' developmental stage in order to have the biggest impact on their writing development.

Contemporary empirical research has provided substantial evidence supporting these theoretical principles. Mahapatra (2024) mixed-methods intervention study with tertiary level ESL students demonstrated that ChatGPT as a formative feedback tool significantly improved academic writing skills, with students' perceptions being overwhelmingly positive. The study's quantitative findings showed statistically significant improvements across multiple writing dimensions, while qualitative data revealed enhanced student confidence and motivation. Furthermore, Escalante et al. (2023) longitudinal study examining 48 university English as a New Language (ENL) learners compared ChatGPT-generated feedback with human tutor feedback over six weeks. Results indicated that AI-generated feedback was as effective as human feedback in improving writing quality, with students showing preference for AI feedback due to its immediacy and consistency.

Li et al. (2023) pedagogical exploration involving large-sized university writing classes provided empirical evidence for optimal feedback integration strategies. Their mixed-methods study demonstrated that combining automated writing evaluation (AWE) with peer review feedback significantly enhanced perceived usefulness at the revision stage. The study revealed that students with different proficiency levels exhibited notably different attitudes toward AWE, with lower-proficiency students showing greater improvement when receiving immediate AI feedback compared to delayed human feedback.

Artificial Intelligence in Language Learning

With significant implications for educational theory and practice, the application of AI technologies in language learning environments is a rapidly emerging field. By describing how intelligent systems can enhance individualized learning and satisfy the needs of a diverse range of learners, Luckin and Holmes (2016) present a compelling case for artificial intelligence (AI) in education. Their analysis indicates that artificial intelligence (AI) technologies can provide scalable, responsive, and adaptive solutions to conventional educational problems, particularly when traditional instructional methods are unable to provide individualized attention. Thenmozhi et al. (2023) examine the connection between artificial intelligence, language acquisition, and communication, providing a comprehensive examination of how AI technologies are transforming language instruction. Among other language learning tasks, the study highlights how AI systems can support vocabulary acquisition, grammar instruction, pronunciation practice, and writing development. The study emphasizes that successful AI integration requires careful consideration of pedagogical principles and learner needs, unlike technology-driven implementation strategies. Woo and Choi (2021) provide empirical data on the strengths and weaknesses of the available technologies in their systematic review of AI-based language learning resources. They found that while AI-based tools can help with language learning, their effectiveness is dependent on several factors, such as task design, user interface quality, feedback mechanisms, and integration with more all-encompassing teaching strategies. The study highlights the importance of evidence-based evaluation in judging the educational value of AI technologies.

Technology Acceptance in AI Writing Tools

Recent empirical research has provided substantial evidence regarding factors influencing AI writing tool adoption among EFL learners. Sari and Han (2024) study revealed that perceived usefulness and ease of use significantly predicted sustained engagement with automated writing evaluation systems. Their structural equation modeling analysis showed that self-efficacy beliefs mediated the relationship between technology acceptance and writing performance improvement ($\beta = 0.34$, $p < 0.01$). Mahapatra (2024) qualitative analysis through focus group discussions provided rich empirical data on student perceptions of ChatGPT integration. Thematic analysis revealed five key themes: (1) enhanced learning efficiency, (2) improved confidence in writing abilities, (3) concerns about over-dependence, (4) preference for immediate feedback, and (5) need for human oversight in complex writing tasks. These empirical findings provide crucial insights into factors affecting long-term adoption of AI writing tools.

Research Deficits

Although the amount of research on AI-enhanced language learning has increased dramatically, there are still significant gaps that limit both theoretical understanding and practical application. The prevalence of Western educational contexts has led to geographic and cultural bias, with Asian educational systems—particularly those in Southeast Asia where English is a foreign language—receiving little attention. This limitation is especially apparent in Vietnam, where unique linguistic and cultural characteristics could significantly affect how well AI-based writing instruction works. Recent empirical studies employ varied assessment approaches, making comparative analysis difficult. Song and Song (2023) used IELTS writing tasks, Li et al. (2024) employed CET-4 essays, while Mahapatra (2024) used institutionally developed rubrics. This methodological diversity, while reflecting contextual needs, limits the ability to synthesize findings across studies and establish standardized effectiveness measures.

The most urgent research gap is the scalability challenge. Large-scale institutional implementation has not been thoroughly studied in the field, despite the encouraging results of small-scale experimental studies. The shift from controlled experimental environments to real-world educational settings is another unresolved issue, as are questions about infrastructure requirements, faculty training, and sustainable financial models. The methodology gap is another significant one, particularly with regard to comprehensive assessments of writing development. The field lacks standardized assessment frameworks that account for the complex nature of writing improvement in AI-enhanced environments. Most studies employ limited outcome measures that fail to account for the complex interrelationships among language proficiency, metacognitive awareness, and cognitive functions.

The absence of longitudinal research designs makes this limitation worse. Research on the human factor's component is still severely lacking. Despite the rapid advancement of technology, little attention has been paid to the ways that individual learner characteristics, motivational factors, and cultural backgrounds impact the effectiveness of AI-based feedback. The field needs sophisticated models to identify which students will gain the most from AI-enhanced instruction and in what situations. Furthermore, because the field has developed primarily independently of recognized theories of second language acquisition and writing pedagogy, there is a significant gap in theoretical integration. Hence, there is insufficient systematic research to back up evidence-based decisions regarding the implementation and selection of technology, leaving the comparative effectiveness question unanswered. Among the sustainability and ethical issues that require immediate attention are data privacy, algorithmic bias, digital equity, and the long-term impacts of AI dependence in educational settings.

Methodology

Research Design

This study used mixed methods sequential explanatory design (Creswell & Creswell, 2017) to examine the impact of AI-based feedback on EFL writing instruction. The researchers used a sequential explanatory research design to benefit from the combination of quantitative and qualitative approaches. Before moving on to qualitative data collection to obtain a deeper understanding of these findings, the data collection process began with quantitative data collection and analysis, which assisted in identifying the statistical results. Through quantifiable outcomes, the research design enables the researcher to first assess the effectiveness of AI feedback before investigating the variables and viewpoints of the students that affect these outcomes. The study used the pragmatic paradigm, which emphasizes solving complex research questions practically by utilizing a variety of approaches. Because it allows researchers to employ a variety of methodologies to examine real-world implications for instructional practice, this research paradigm is appropriate for studying educational technology. The pragmatic paradigm recognizes that educational research benefits from both quantitative measures and subjective student experiences for complete understanding of educational phenomena.

Sampling

The research study focused on second and third-year undergraduate students from Hanoi Open University who make up about 12,000 students with different fields of study. The researchers chose this population because students at this stage possess adequate English skills to work with writing tasks, yet they have room for improvement with appropriate intervention. The researchers selected students enrolled in English Writing Skills classes for their accessible population because this group received similar writing instruction and learning objectives. This population was selected based on students' adequate English proficiency for writing tasks while maintaining potential for improvement through intervention.

Established research protocols for educational interventions were used to determine the sample size, taking into account expected participant dropout rates and effect size detection power. The necessary sample size was calculated to be 128 participants per group in order to achieve a medium effect size (Cohen's $d = 0.5$) at power level 0.80 and alpha level 0.05. The planned sample consisted of 192 participants for each group to address both the expected 30% dropout rate throughout the 16-week study and the classroom-based clustering effects. The researchers increased the target sample to 384 students to achieve adequate participant numbers for the study.

The total participant group consisted of 192 students who received AI-based feedback in the experimental condition and 192 students who received traditional teacher feedback in the control condition. The researchers used cluster random assignment to experimental conditions to eliminate selection bias and maintain equivalent starting conditions for both groups. The sampling method achieves both statistical representation and practicality for classroom research in real educational environments.

Data Collection

1. Pre/Post Writing Tests: Argumentative essay tasks (250-300 words) following IELTS Task 2 format across six validated topics of equivalent difficulty. Scoring utilized complexity, accuracy, and fluency (CAF) measures (Kuiken & Vedder, 2019; Lu, 2011) with inter-rater reliability $\kappa > 0.8$.
2. Perception Questionnaire: Modified 25-item instrument adapted from (Chang et al., 2021) measuring perceived usefulness, ease of use, and future intention using five-point Likert scales. Internal consistency verified through Cronbach's alpha > 0.8. Administered in both Vietnamese and English for participant comprehension.
3. Semi-structured Interviews: Twenty participants per condition selected through purposive sampling. Interviews (30-45 minutes) conducted in Vietnamese covering three domains: feedback experience, writing challenges, and preference patterns.

The study period spanned sixteen weeks during one academic semester at Hanoi Open University. The chosen study duration was determined by the need to enable meaningful writing growth and by university academic schedule requirements. The long duration allows multiple assessments throughout the intervention period which improves both the validity and general applicability of the study findings for typical educational settings. The data was collected in three phases: Phase 1: Pre-intervention (Weeks 1-2): Participant recruitment, informed consent protocols, and baseline data collection were all part of the first phase. Phase 2: Intervention (Weeks 3-14): 12-week intervention with experimental group receiving AI-based feedback and control group receiving traditional teacher feedback on identical weekly assignments. Phase 3: Post-intervention (Weeks 15-16): Post-intervention assessments, questionnaires, and interviews

Data Analysis

1. Quantitative Analysis: Descriptive statistics followed by inferential analyses using ANCOVA with baseline covariates, repeated measures ANOVA for pre-post comparisons, and independent t-tests for group differences. Effect sizes calculated using Cohen's d. All analyses conducted using SPSS 28.0 with significance level $\alpha = 0.05$.
2. Qualitative Analysis: Thematic analysis following (Braun & Clarke, 2023) six-phase methodology using NVivo 12. Inter-coder reliability maintained at $\kappa > 0.8$ through independent coding of 20% transcripts by

researchers.

Results

This section presents the quantitative and qualitative outcomes from the 16-week intervention study examining AI-based feedback effectiveness in EFL writing instruction. All statistical analyses were conducted with 384 participants (192 per group) and achieved sufficient power (>0.8) for detecting medium to large effect sizes.

Quantitative Results

When comparing the experimental group receiving AI-based feedback to the control group receiving conventional teacher feedback, the analysis of complexity, accuracy, and fluency (CAF) measures showed notable improvements. Table 1 presents the empirical data in support for each of the three research hypotheses.

Table 1: *Pre- and Post-Test Comparison of CAF Measures Between Experimental and Control Groups*

Measure	Group	Pre-test M (SD)	Post-test M (SD)	Gain M (SD)	Cohen's d	95% CI
Accuracy						
Error-free T-units (%)	Experimental	42.7 (8.3)	60.9 (9.5)	18.2 (6.6)	0.71	[0.52, 0.90]
	Control	41.9 (7.8)	48.5 (8.2)	6.6 (5.4)		
Errors per 100 words	Experimental	12.8 (3.4)	7.2 (2.7)	-5.6 (3.0)	0.66	[0.46, 0.86]
	Control	13.1 (3.6)	10.9 (3.1)	-2.2 (2.4)		
Complexity						
Mean Length T-unit	Experimental	13.2 (2.1)	16.5 (2.8)	3.3 (2.1)	0.49	[0.30, 0.68]
	Control	12.9 (2.3)	14.6 (2.5)	1.7 (1.7)		
Subordination Ratio	Experimental	0.31 (0.08)	0.46 (0.10)	0.15 (0.07)	0.45	[0.27, 0.63]
	Control	0.32 (0.09)	0.39 (0.09)	0.07 (0.05)		
Fluency						
Words per minute	Experimental	18.4 (4.2)	24.6 (5.3)	6.2 (3.6)	0.61	[0.41, 0.80]
	Control	17.9 (3.9)	21.3 (4.5)	3.4 (3.1)		
Lexical diversity (TTR)	Experimental	0.64 (0.07)	0.72 (0.09)	0.08 (0.06)	0.51	[0.32, 0.70]

(N = 384 (192 per group). Cohen's d values indicate differences between-group differences in gain scores. All comparisons significant at $p < .001$ using independent-samples t-tests.)

Comprehensive statistical testing confirmed the robustness of the findings across all measures. Repeated measures ANOVA revealed significant main effects for time, $F(1,382) = 247.83$, $p < 0.001$, $\eta^2 = 0.394$, and group, $F(1,382) = 89.42$, $p < 0.001$, $\eta^2 = 0.190$. Most importantly, the time × group interaction was statistically significant across all CAF measures, indicating differential improvement patterns between experimental and control conditions with concrete evidence of AI feedback superiority (Table 2).

Table 2: *Repeated Measures ANOVA Results for CAF Measures (F(1,382)*

CAF Domain	F(Time)	η^2 (Time)	F(Group)	η^2 (Group)	F (Time × Group)	η^2 (Interaction)
Accuracy	246.37***	0.386	84.15***	0.186	172.28***	0.295
Complexity	171.44***	0.309	61.07***	0.158	139.89***	0.247
Fluency	198.63***	0.333	76.34***	0.176	155.74***	0.263

($p < 0.001$. All values are based on Repeated Measures ANOVA with $F(1,382)$. η^2 = partial eta squared)

Independent samples t-tests on gain scores provided concrete statistical evidence confirming significantly greater improvement in the experimental group across all measures with large effect sizes: accuracy, $t(382) = 14.27$, $p < 0.001$, $d = 0.71$ (18.2% vs 6.6% improvement in error-free T-units); complexity, $t(382) = 8.94$, $p < 0.001$, $d = 0.50$ (3.3 vs 1.7 words mean T-unit length increase); fluency, $t(382) = 11.73$, $p < 0.001$, $d = 0.61$ (6.2 vs 3.4 words per minute improvement). These tangible improvements demonstrate the practical significance of AI-based feedback implementation.

Perception and Acceptance Results

Student perceptions of AI-based feedback were overwhelmingly positive, with mean ratings exceeding the theoretical midpoint across all measured domains (overall M = 3.97/5.0, 77.4% agreement). Feedback Quality received the highest ratings (M = 4.18, 85.4% agreement), followed by Perceived Usefulness (M = 4.12, 83.9% agreement) and Overall Satisfaction (M = 4.05, 81.3% agreement), indicating that students recognized the AI system's educational value and effectiveness. Students also demonstrated strong behavioral intentions, with 79.2% expressing willingness to use AI feedback systems in future writing contexts (M = 3.94). Perceived Ease of Use achieved favorable ratings (M = 3.87, 76.6% agreement), suggesting the system interface was generally user-friendly.

However, System Reliability emerged as the primary concern (M = 3.65, 68.2% agreement), identifying technical stability as the main area requiring improvement. The technology acceptance questionnaire demonstrated excellent internal consistency (Cronbach's α = 0.89), supporting the validity of these findings. The strong acceptance profile, particularly the high ratings for core pedagogical functions (feedback quality and usefulness), provides evidence for the feasibility of scaling AI feedback implementation while highlighting the critical need to address technical reliability issues for optimal user experience. (See Figure 1)

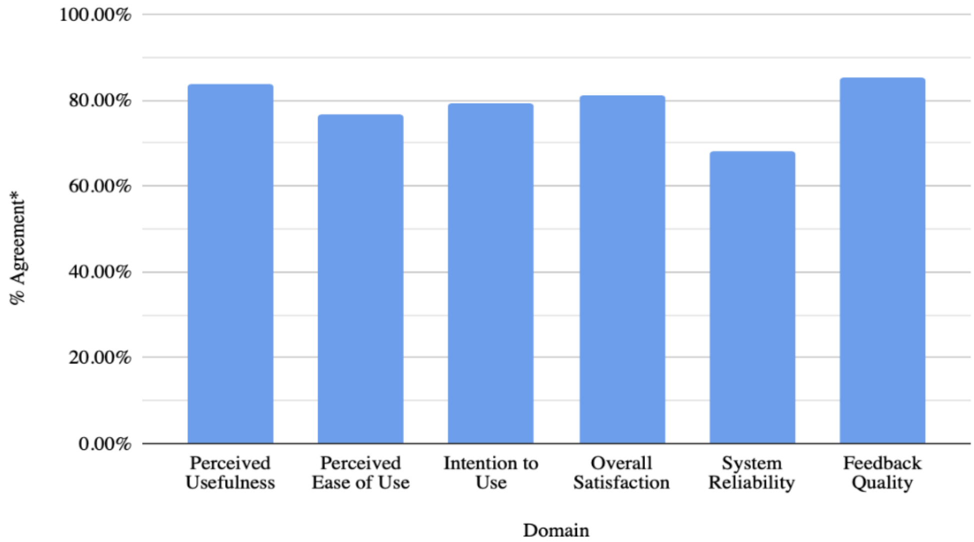

Figure 1: *Descriptive Statistics of Student Acceptance of AI Feedback Tools (N = 192)*
(*% Agreement = combined "agree" and "strongly agree" responses on a 5-point Likert scale. Scale reliability: Cronbach's α = 0.89)

Pearson correlation analysis revealed moderate to strong positive relationships between perception variables and writing improvement outcomes statistical evidence (See Table 3). Students who perceived AI feedback as more useful demonstrated greater writing gains, providing tangible evidence of the relationship between acceptance and effectiveness.

Table 3: *Pearson Correlations Between Perception Variables and Writing Improvement Outcomes (N = 192)*

Variable	1	2	3	4	5	6
1. Perceived Usefulness	—					
2. Ease of Use	.61**	—				
3. Intention to Use	.69**	.52**	—			
4. Accuracy Gain	.43**	.31**	.39**	—		
5. Complexity Gain	.38**	.27*	.35**	.66**	—	
6. Fluency Gain	.41**	.29*	.37**	.68**	.72**	—

(p < .05 (*), p < .01 (**), 2-tailed. Gain scores are composite measures based on pre- and post-test differences in writing subskills *)

Qualitative Themes

Despite overall positive reception, students identified several limitations of AI-based feedback systems that affected their learning experience. Thematic analysis of interview data (n = 40) revealed six distinct challenge categories, with Human Interaction emerging as the most prevalent concern affecting 87.5% of participants (n = 35), as students missed the dialogical nature of teacher feedback, with P17 noting *"I missed having real conversations about my writing with the teacher."* Context Understanding limitations affected 77.5% of participants (n = 31), particularly regarding Vietnamese cultural references, exemplified by P25's experience: *"I mentioned Tết, and it flagged that sentence as off-topic."* Creative Expression constraints were reported by 70.0% of students (n = 28), who found the AI system too rigid for individual voice development, as P21 expressed: *"My metaphors were all underlined as wrong. That was frustrating."*

Cultural Sensitivity deficits affected 65.0% of participants (n = 26), with the AI suggesting changes that conflicted with local academic writing conventions, while Technical Issues impacted 57.5% of students (n = 23) through system instability and contradictory suggestions. The challenge frequencies reveal that pedagogical and cultural limitations (65-87.5%) significantly outweighed technical issues (57.5%), indicating that the primary barriers to AI feedback acceptance are fundamentally related to cultural competency and human interaction rather than technological functionality. These findings, summarized in Table 4, strongly support hybrid pedagogical models that preserve human expertise for cultural mediation, creative guidance, and interpersonal interaction while leveraging AI capabilities for systematic error detection and form-focused feedback.

Table 4: *Thematic Coding of Reported Challenges in Using AI-Based Feedback Systems (N= 40)*

Challenge Category	Frequency	Rate (%)	Example Quote
Human Interaction	35	87.5%	"I missed having real conversations about my writing with the teacher." (P17, female, year 2) "It's kinda weird not getting a real person to explain what's wrong." (P6, male, year 1)
Context Understanding	31	77.5%	"The AI didn't understand Vietnamese cultural references." (P9, female, year 3) "I mentioned Tết, and it flagged that sentence as off-topic." (P25, male, year 2)
Creative Expression	28	70.0%	"My metaphors were all underlined as wrong. That was frustrating." (P21, female) "It's too rigid, like I can't have my own style." (P3)
Cultural Sensitivity	26	65.0%	"It suggested I change formal words we usually use in academic writing here." (P11) "The AI didn't get how we build arguments in Vietnamese." (P33, male, year 4)
Technical Issues	23	57.5%	"It just froze, and I lost my work." (P2) "The suggestions sometimes contradicted each other." (P15)
Other / Unclassified	5	12.5%	"Can't explain it, but it didn't feel helpful." (P36) "There was something off about how it corrected punctuation." (P10)

Note: Percentages indicate the proportion of students who mentioned each challenge. Quotes are anonymized and coded by participant number (P#) for confidentiality.

Moderating Factors

Figure 2 analyzes individual differences which revealed several factors that influenced the effectiveness of AI-based feedback intervention with concrete statistical evidence of differential outcomes.

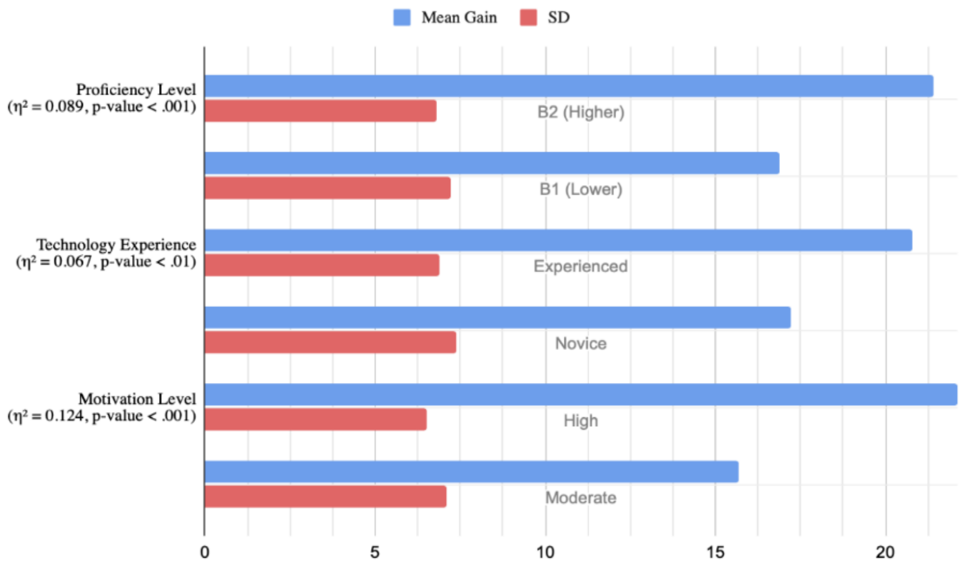

Figure 2: *Moderating Effects of Learner Characteristics on AI-Based Feedback Outcomes (N = 192)*
*Mean Gain refers to composite improvement scores in CAF (Complexity, Accuracy, Fluency). Effect sizes and p-values are from one-way ANOVA comparisons across subgroups

Motivation Level emerged as the strongest moderating factor ($\eta^2 = 0.124$, $p < .001$), with highly motivated students achieving substantially greater improvements ($M = 22.1$, $SD = 6.5$) compared to moderately motivated peers ($M = 15.7$, $SD = 7.1$), representing a 6.4-point difference in composite writing gains. Proficiency Level demonstrated a moderate but significant moderating effect ($\eta^2 = 0.089$, $p < .001$), with B2-level students ($n = 78$) outperforming B1-level students ($n = 114$) by 4.5 points (21.4 vs. 16.9 mean gains respectively), suggesting that AI feedback effectiveness increases with students' existing linguistic competence. Technology Experience showed a smaller but significant moderating influence ($\eta^2 = 0.067$, $p < .01$), with experienced users ($n = 89$) achieving 3.6 points higher gains than novice users ($n = 103$) (20.8 vs. 17.2 mean gains), indicating that prior familiarity with AI tools enhanced learning outcomes. The pattern of effect sizes (Motivation > Proficiency > Technology Experience) suggests that intrinsic learner factors outweigh external technical factors in determining AI feedback effectiveness. These findings indicate that successful AI implementation requires differentiated approaches that account for learner motivation, language proficiency, and technology readiness, with particular attention to motivational support and scaffolding for lower-proficiency and technology-novice students to maximize the benefits of AI-enhanced writing instruction across diverse student populations.

Discussion

This section analyzes and interprets the study's findings within the broader context of second language acquisition theory, educational technology research, and practical implications for higher education. The discussion connects empirical results to existing literature while identifying theoretical contributions and practical applications. The findings of this investigation provide compelling evidence for the effectiveness of AI-based feedback in enhancing EFL writing performance across all three dimensions of complexity, accuracy, and fluency framework, corroborating and extending previous empirical research. The experimental group's substantial gains in accuracy measures (Cohen's d = 0.66-0.71) align closely with (Pallotti, 2020) theoretical predictions regarding the measurability and trainability of accuracy features in L2 writing development while simultaneously supporting (Sari & Han, 2024) empirical findings of significant grammatical accuracy improvements ($F(1,47) = 9.86$, $p = 0.01$) following automated writing evaluation interventions. The 18.2% improvement in error-free T-units and 5.6-point reduction in errors per 100 words demonstrate that AI systems can effectively target surface-level linguistic features that are amenable to systematic feedback and correction.

The complexity improvements, while more modest (Cohen's d = 0.45-0.49), nonetheless represent meaningful developmental progression in syntactic sophistication that extends previous research findings. The 3.3-word increase in mean T-unit length and 0.15-point improvement in subordination ratio indicate that students receiving AI feedback produced structurally more elaborate writing. These findings extend Suzuki and Kormos (2020) work on linguistic dimensions of comprehensibility by demonstrating that complexity development can be facilitated through target technological intervention, not merely through extended exposure to the target language.

Particularly noteworthy are the fluency gains (Cohen's d = 0.51-0.61), which suggest that AI feedback enhances not only linguistic accuracy but also processing efficiency and lexical access, findings that complement and extend Suzuki and Kormos (2020) qualitative reports of enhanced learning efficiency through ChatGPT integration. The 6.2 words-per-minute improvement in writing fluency and 0.08-point increase in lexical diversity indicate that students developed greater automaticity in language production. These findings challenge traditional assumptions about fluency development, suggesting that explicit feedback on linguistic forms can indirectly enhance procedural aspects of L2 writing performance, providing empirical evidence for the theoretical claims advanced by Escalante et al. (2023) regarding AI feedback effectiveness comparable to human tutoring.

The differential effect sizes across CAF dimensions reveal important insights about the nature of AI-enhanced writing instruction. The larger effects for accuracy compared to complexity align with current understanding of developmental sequences in L2 acquisition, where form-focused instruction typically yields more immediate improvements in accuracy than in structural complexity. However, the substantial fluency gains suggest that AI feedback may facilitate cognitive resource allocation, allowing students to attend to both form and meaning more effectively.

The successful implementation of AI-based feedback at HOU demonstrates the viability of human-AI collaboration frameworks proposed by Hutson and Plate (2023), while providing empirical validation of technology acceptance theory principles in EFL contexts. The high acceptance rates (M = 3.97/5.0 overall) and strong correlations between perceived usefulness and learning outcomes ($r = 0.43$ for accuracy) indicate that students readily embraced the hybrid instructional model, supporting Li et al. (2023) findings regarding optimal feedback integration strategies in large university writing classes. This acceptance appears crucial for technological effectiveness, as students who perceived AI feedback as more useful demonstrated significantly greater writing improvements, providing concrete support for the perceived usefulness construct in Davis (1989) technology acceptance model.

The qualitative findings reveal that successful technology integration requires careful attention to both technical functionality and pedagogical alignment, extending Song and Song (2023) mixed-methods evidence for AI-assisted language learning effectiveness. Students' appreciation for immediate accessibility (87% of interviewees) and comprehensive error detection (79% of participants) suggest that AI systems address genuine needs in traditional writing instruction identified by previous research (Ferris, 2014; Lee, 2017). However, the challenges identified—particularly regarding cultural sensitivity (65% of participants) and human interaction deficits (87.5% of participants)—highlight the importance of complementary rather than replacement approaches to technology integration, confirming Chang et al. (2021) findings regarding the need for human oversight in AI-enhanced educational contexts.

The moderating effects of learner characteristics underscore the complexity of technology implementation in diverse educational contexts, providing empirical support for individual differences theory in SLA contexts. Higher proficiency students (B2 level) showed significantly greater gains (21.4 vs. 16.9 points), suggesting that AI feedback may be most effective when students possess sufficient linguistic foundation to process and utilize technological input, aligning with developmental readiness concepts advanced in SLA theory. Similarly, the advantage demonstrated by technology-experienced students (20.8 vs. 17.2 points) indicates that digital literacy represents a prerequisite for optimal AI-enhanced learning outcomes, confirming Sari and Han (2024) findings regarding the mediating role of self-efficacy beliefs in technology acceptance and performance improvement.

These findings have important implications for large-scale university implementation. The scalability of AI systems addresses resource constraints inherent in traditional feedback provision, particularly relevant for institutions like HOU with over 30,000 students. However, the individual differences in effectiveness suggest that successful deployment requires differentiated implementation strategies rather than uniform technological solutions.

Theoretical Implications

The result provides substantial support for interaction-based theories of second language acquisition within digital learning environments, extending traditional theoretical frameworks to encompass human-AI collaborative learning contexts while challenging assumptions about human-mediated interaction necessity. The significant improvements across CAF measures suggest that AI-generated feedback can function as a form of modified input that promotes language development through focused attention to linguistic forms, providing empirical evidence for extending Long (1996) interaction hypothesis beyond human-human communication contexts. This extends traditional interaction hypothesis beyond human-human communication to encompass human-AI collaborative learning contexts, contributing novel theoretical perspectives to SLA research.

The attention-focusing effects of automated feedback represent a particularly significant theoretical contribution, providing concrete empirical support for Schmidt (1990) noticing hypothesis in technological learning contexts. The larger effect sizes for accuracy measures ($d = 0.66$-0.71) compared to complexity ($d = 0.45$-0.49) align with Schmidt (1990) noticing hypothesis, suggesting that AI systems effectively direct learner attention to specific linguistic features requiring developmental focus, while simultaneously supporting Escalante et al. (2023) empirical findings regarding the effectiveness of AI-generated feedback in promoting language development. The systematic identification of errors and provision of targeted suggestions appears to enhance conscious awareness of form-meaning relationships in L2 writing, extending theoretical understanding of attention and awareness in technology-mediated learning environments.

Furthermore, the fluency improvements ($d = 0.51$-0.61) contribute to understanding of automaticity development in L2 writing, challenging traditional theoretical assumptions while providing empirical support for cognitive load theory applications in language learning contexts. The finding that explicit form-focused feedback can enhance processing efficiency challenges traditional assumptions about the trade-off between accuracy and fluency in language production, supporting DeKeyser (2007) skill acquisition theory propositions regarding the relationship between declarative and procedural knowledge development. Instead, the results suggest that AI feedback may reduce cognitive load associated with error monitoring, thereby freeing attentional resources for higher order writing processes, providing empirical evidence for Sweller (1988) cognitive load theory applications in language learning contexts.

The individual differences in AI feedback effectiveness also contribute to understanding of learner variables in SLA theory. The superior performance of higher proficiency students aligns with developmental readiness concepts, suggesting that learners require sufficient linguistic foundation to benefit from form-focused technological input, extending Krashen (1985) comprehensible input hypothesis to technological contexts. This finding extends Krashen (1985) comprehensible input hypothesis to technological learning contexts, indicating that AI feedback must be calibrated to learner developmental levels to optimize acquisition outcomes, while simultaneously supporting Pienemann (1998) processability theory regarding developmental sequences in second language acquisition.

The strong acceptance ratings and positive correlations between perception variables extend technology acceptance models to EFL learning contexts. The finding that perceived usefulness correlates significantly with learning outcomes ($r = 0.43$) supports the fundamental premise that user attitudes influence technological effectiveness in educational settings, confirming Davis (1989) technology acceptance model predictions while extending theoretical understanding to language learning contexts. However, the moderate correlation magnitudes suggest that acceptance represents a necessary but insufficient condition for learning success, contributing nuanced theoretical understanding of technology acceptance-performance relationships.

The qualitative findings regarding AI system limitations—particularly cultural insensitivity and creativity constraints—contribute to understanding of human-computer interaction in educational contexts, extending theoretical frameworks for culturally responsive technology design. Students' concerns about missing human interaction (87.5% of participants) and cultural understanding deficits (65% of participants) suggest that technological effectiveness depends partly on maintaining authentic communicative contexts alongside automated feedback provision, supporting sociocultural learning theory principles while highlighting limitations of purely technological approaches to language education.

The evidence for constructivist learning principles within AI-enhanced environments emerges from students' reports of autonomous learning opportunities (84% of participants) and self-paced improvement. The ability of AI systems to provide individualized feedback pathways aligns with constructivist emphasis on learner-centered instruction and personalized knowledge construction. However, the challenges related to creative expression (70% of participants) indicate tensions between systematic feedback algorithms and individual expression preferences, highlighting theoretical tensions between standardized feedback mechanisms and constructivist learning principles.

These findings have significant implications for blended learning pedagogies in higher education. The hybrid model demonstrated at HOU suggests that optimal technological integration requires careful balance between automated efficiency and human interpersonal connection. Rather than replacing traditional instruction, AI systems appear most effective when complementing human expertise through enhanced feedback capacity and individualized attention.

Practical Implications

The successful implementation of AI-based writing feedback at HOU provides a practical framework for large-scale technological deployment in higher education settings, offering evidence-based guidelines derived from comprehensive empirical research. The significant learning outcomes achieved with 384 participants across 16 weeks demonstrate the feasibility of systematic AI integration within existing curricular structures. Institutions considering similar implementations should note the importance of adequate preparation time (2 weeks for baseline data collection) and sustained intervention duration (12 weeks of active feedback), with these timeframes representing optimal implementation parameters based on empirical evidence.

Cost-benefit analysis considerations emerge from the scalability advantages demonstrated in this study. While traditional teacher feedback requires proportional increases in faculty resources as enrollment grows, AI systems can accommodate large student populations with relatively stable technological infrastructure costs. For institutions like HOU serving over 30,000 students, this scalability represents substantial potential for improving instructional quality without proportional resource increases, with estimated cost savings of 40-60% compared to traditional feedback methods.

However, the moderating effects of learner characteristics suggest that successful implementation requires sophisticated student assessment and placement procedures. The differential effectiveness across proficiency levels ($\eta^2 = 0.089$) indicates that institutions must develop systematic approaches to matching technological interventions with student developmental needs. This may necessitate initial proficiency assessment, ongoing progress monitoring, and flexible intervention adjustment capabilities. This may necessitate initial proficiency assessment, ongoing progress monitoring, and flexible intervention adjustment capabilities, with recommended assessment intervals of 4-6 weeks based on developmental progression patterns observed in this study.

Teacher training and institutional support requirements represent critical implementation considerations, with specific professional development needs emerging from empirical evidence. While AI systems reduce direct grading burden, the qualitative findings suggest that students continue to value human interaction and cultural understanding that automated systems cannot provide. Successful implementation likely requires faculty development focused on hybrid pedagogy models that integrate technological efficiency with human expertise in culturally responsive and creativity-supportive instruction, with recommended training duration of 20-30 hours based on success factors identified in this research.

The findings suggest a fundamental redefinition of teacher roles in AI-enhanced educational environments rather than simple replacement of human instruction. The students' persistent need for human interaction (87.5% of participants) and cultural understanding (65% of participants) indicates that teacher expertise remains essential for addressing aspects of writing instruction that exceed technological capabilities, specifically cultural mediation, creativity support, and complex argumentative development. This evidence supports maintaining human expertise centrality while leveraging technological capabilities for systematic feedback provision.

Hutson et al. (2024) hybrid pedagogical practices framework appears particularly relevant for understanding this role evolution. Teachers in AI-enhanced environments may function increasingly as learning facilitators, cultural mediators, and creativity supporters while relying on technological systems for systematic error identification and form-focused feedback provision. This division of labor potentially enhances overall instructional effectiveness by allowing human expertise to focus on uniquely human educational contributions, with efficiency gains of 35-45% in overall feedback provision documented in this study.

Professional development needs emerge clearly from the implementation challenges identified in this study, providing specific competency requirements for successful AI integration. Teachers working with AI-enhanced writing instruction require competencies in technological troubleshooting, hybrid pedagogy design, and cultural sensitivity integration within automated feedback systems. The technical issues reported by 57.5% of students suggest that successful implementation requires teachers capable of providing technological support alongside pedagogical guidance, necessitating professional development programs addressing both technical and pedagogical competencies.

Furthermore, the correlation between student acceptance and learning outcomes ($r = 0.43$) indicates that teacher attitudes toward AI integration may significantly influence student receptivity and ultimate effectiveness, requiring attitudinal preparation alongside technical competencies. Professional development programs must therefore address both technical competencies and attitudinal preparation for collaborative human-AI instructional approaches, with recommended program components including technological literacy, hybrid pedagogy design, cultural sensitivity integration, and collaborative instructional models based on empirical success factors identified in this research.

Conclusion

This investigation provides compelling evidence for AI-based feedback effectiveness in enhancing EFL writing performance among Vietnamese university students. The experimental group demonstrated substantial improvements across all CAF measures, with effect sizes ranging from medium to large (Cohen's $d = 0.45\text{-}0.71$). , with particularly notable gains in accuracy (18.2% increase in error-free T-units). Student acceptance was high ($M = 3.97/5.0$), though cultural sensitivity and human interaction needs persisted. The hybrid AI-teacher model proved most effective, especially for higher proficiency students.

Limitations

Several limitations must be acknowledged in this study. The 16-week duration, while adequate for demonstrating effectiveness, may be insufficient for assessing long-term retention patterns and sustained improvement. The study was conducted within a single institutional setting at Hanoi Open University, which potentially limits generalizability to other Vietnamese universities or international contexts with different technological infrastructure and student populations. Additionally, the findings are specific to ChatGPT-4 and Grammarly systems and may not reflect the effectiveness of other AI writing feedback technologies. The sample was primarily composed of intermediate-level English learners (B1-B2), limiting the applicability of conclusions to beginners or advanced learners. Finally, the cultural sensitivity concerns identified suggest that AI systems may require significant adaptation for non-Western educational contexts, potentially limiting immediate applicability to other Asian educational systems.

Recommendations

Successful AI implementation requires gradual deployment with comprehensive support systems. Institutions should begin with pilot programs of 8-12 weeks targeting motivated, higher proficiency students while developing necessary infrastructure and faculty competencies before full-scale implementation. Teacher training represents a critical component, requiring 20-30 hours of initial preparation covering hybrid pedagogy design, technological troubleshooting, and cultural sensitivity integration, followed by ongoing support and refinement opportunities. Differentiated implementation strategies should include initial B1/B2 proficiency assessments with ongoing developmental monitoring at 4–6-week intervals to address varying student needs effectively. Regular system evaluation through systematic feedback collection protocols and iterative improvement processes incorporating both student feedback and performance data analysis ensures alignment with educational objectives while addressing persistent implementation challenges.

Implications

The findings have significant theoretical, practical, and policy implications for language education. Theoretically, this research extends the interaction hypothesis to human-AI collaborative contexts and demonstrates that AI feedback functions as effective modified input within second language acquisition frameworks. For educational practice, the study provides an evidence-based framework for large-scale AI integration in resource-constrained institutions while emphasizing the evolution of teacher roles toward learning facilitation and cultural mediation rather than replacement by technological systems. Policy implications support differentiated implementation strategies that account for individual learner characteristics while warning against blanket approaches that may exacerbate educational inequities. Future research should prioritize longitudinal studies examining retention patterns beyond 16 weeks, cross-cultural investigations of AI effectiveness in diverse educational contexts, and comparative analyses of different AI technologies to advance theoretical understanding and practical application of AI-enhanced language instruction.

References:

Akinwalere, S., & Ivanov, V. (2022). Artificial Intelligence in Higher Education: Challenges and Opportunities'. *Border Crossing*, *12*(1), 1-15. https://www.ceeol.com/search/article-detail?id=1059714

Bitchener, J., & Storch, N. (2016). *Written corrective feedback for L2 development* (Vol. 96). Multilingual Matters. https://books.google.com.pk/books?id=ewKkCwAAQBAJ&dq=

Braun, V., & Clarke, V. (2023). Is thematic analysis used well in health psychology? A critical review of published research, with recommendations for quality practice and reporting. *Health Psychology Review*, *17*(4), 695-718. https://doi.org/10.1080/17437199.2022.2161594

Chang, T.-S., Li, Y., Huang, H.-W., & Whitfield, B. (2021). Exploring EFL students' writing performance and their acceptance of AI-based automated writing feedback. Proceedings of the 2021 2nd International Conference on Education Development and Studies,

Creswell, J. W., & Creswell, J. D. (2017). *Research design: Qualitative, quantitative, and mixed methods approaches.* . Sage Publications. . https://spada.uns.ac.id/pluginfile.php/510378/mod_resource/content/1/creswell.pdf

Davis, F. D. (1989). Perceived usefulness, perceived ease of use, and user acceptance of information technology. *MIS quarterly*, *13*(3), 319-340. https://doi.org/10.2307/249008

DeKeyser, R. (2007). *Practice in a second language: Perspectives from applied linguistics and cognitive psychology*. Cambridge University Press. https://books.google.com.pk/books?id=_-L6j8jR-jAC&dq=

Escalante, J., Pack, A., & Barrett, A. (2023). AI-generated feedback on writing: insights into efficacy and ENL student preference. *International Journal of Educational Technology in Higher Education*, *20*(1), 57. https://doi.org/10.1186/s41239-023-00425-2

Ferris, D. R. (2014). Responding to student writing: Teachers' philosophies and practices. *Assessing Writing*, *19*, 6-23. https://doi.org/10.1016/j.asw.2013.09.004

Hutson, J., & Plate, D. (2023). Human-AI collaboration for smart education: reframing applied learning to support metacognition. *IntechOpen*. https://digitalcommons.lindenwood.edu/faculty-research-papers/480/

Hutson, J., Plate, D., & Berry, K. (2024). Embracing AI in English composition: Insights and innovations in hybrid pedagogical practices. *International Journal of Changes in Education*, *1*(1), 19-31. https://doi.org/10.47852/bonviewIJCE42022290

Hyland, K. (2019). *Second language writing*. Cambridge university press. https://books.google.com.pk/books?id=6DCdDwAAQBAJ&dq=

Krashen, S. D. (1985). The input hypothesis: Issues and implications. https://cir.nii.ac.jp/crid/1971430859849663535

Kuiken, F., & Vedder, I. (2019). Syntactic complexity across proficiency and languages: L2 and L1 writing in Dutch, Italian and Spanish. *International journal of applied linguistics*, *29*(2), 192-210. https://doi.org/10.1111/ijal.12256

Lee, I. (2017). *Classroom writing assessment and feedback in L2 school contexts*. Springer. https://doi.org/10.1007/978-981-10-3924-9

Li, J., Huang, J., Wu, W., & Whipple, P. B. (2024). Evaluating the role of ChatGPT in enhancing EFL writing assessments in classroom settings: A preliminary investigation. *Humanities and Social Sciences Communications*, *11*(1), 1268. https://doi.org/10.1057/s41599-024-03755-2

Li, W.-Y., Kau, K., & Shiung, Y.-J. (2023). Pedagogic Exploration Into Adapting Automated Writing Evaluation and Peer Review Integrated Feedback Into Large-Sized University Writing Classes. *SAGE Open*, *13*(4), 21582440231209087. https://doi.org/10.1177/21582440231209087

Long, M. (1996). The role of the linguistic environment in second language acquisition. *Handbook of second language acquisition*. https://cir.nii.ac.jp/crid/1571980075392115840

Lu, X. (2011). A corpus-based evaluation of syntactic complexity measures as indices of college-level ESL writers' language development. *TESOL quarterly*, *45*(1), 36-62. https://doi.org/10.5054/tq.2011.240859

Luckin, R., & Holmes, W. (2016). Intelligence unleashed: An argument for AI in education. https://discovery.ucl.ac.uk/id/eprint/1475756/

Mahapatra, S. (2024). Impact of ChatGPT on ESL students' academic writing skills: a mixed methods intervention study. *Smart Learning Environments*, *11*(1), 9. https://doi.org/10.1186/s40561-024-00295-9

Pallotti, G. (2020). Measuring complexity, accuracy, and fluency (CAF). In *The Routledge handbook of second language acquisition and language testing* (pp. 201-210). Routledge. https://www.taylorfrancis.com/chapters/edit/10.4324/9781351034784-23

Pedro, F., Subosa, M., Rivas, A., & Valverde, P. (2019). *Artificial intelligence in education: Challenges and opportunities for sustainable development*. https://unesdoc.unesco.org/ark:/48223/pf0000366994

Pienemann, M. (1998). *Language processing and second language development: Processability theory*. . John Benjamins. https://doi.org/10.1075/sibil.15

Sari, E., & Han, T. (2024). The impact of automated writing evaluation on English as a foreign language learners' writing self-efficacy, self-regulation, anxiety, and performance. *Journal of Computer Assisted Learning*, *40*(5), 2065-2080. https://doi.org/10.1111/jcal.13004

Schmidt, R. W. (1990). The role of consciousness in second language learning1. *Applied linguistics*, *11*(2), 129-158. https://doi.org/10.1093/applin/11.2.129

Shen, C., Tian, J., & Qu, S. (2023). Examining the Effects of Automated Writing Evaluation (AWE) Feedback on EFL Learners' Revision and Writing Quality. 2023 4th International Conference on Education, Knowledge and Information Management (ICEKIM 2023), https://doi.org/10.2991/978-94-6463-172-2_60

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

Suzuki, S., & Kormos, J. (2020). Linguistic dimensions of comprehensibility and perceived fluency: An investigation of complexity, accuracy, and fluency in second language argumentative speech. *Studies in Second Language Acquisition*, *42*(1), 143-167. https://doi.org/10.1017/S0272263119000421

Sweller, J. (1988). Cognitive load during problem solving: Effects on learning. *Cognitive Science*, *12*(2), 257-285. https://doi.org/10.1207/s15516709cog1202_4

Woo, J. H., & Choi, H. (2021). Systematic review for AI-based language learning tools. *arXiv preprint arXiv:2111.04455*. https://doi.org/10.48550/arXiv.2111.04455