ChatGPT-Assisted Feedback and the Development of Lexical Resources in EFL Argumentative Writing: Students' Perceptions and Motivation

Thang Ho-Minh[1]

[1] University of Economics Ho Chi Minh City, Vietnam
*Corresponding author's email address: thangemtoi@ueh.edu.vn
[1]ORCID: https://orcid.org/0000-0002-1514-3974
DOI:

Received: 10/11/2025 Revision: 20/08/2026 Accepted: 17/09/2026 Online: 09/11/2026

ABSTRACT

This study looks at the impacts of ChatGPT on EFL argumentative writing using a case study design with 120 Vietnamese EFL learners. Participants were divided into an experimental and a control group. The former was provided with ChatGPT's feedback; the latter, traditional teacher feedback, along with a post-intervention writing assessment and a questionnaire. The study was evaluated based on PLS-SEM. The findings showed that participants in the experimental group showed more growth in lexical resource development than those in the control group. In addition, participants in both groups expressed high levels of perceived usefulness, perceived ease of use, and learning motivation. Through structural model analysis, it was shown that lexical resource development was positively affected by perceived ease of use and perceived usefulness with learning motivation as the mediator. Accordingly, lexical resource development was positively affected by the quality of the AI's feedback and the participants' willingness to use the technology within the given constraints and their motivation to revise. Thus, as the first study of its kind, it confirms that AI feedback has the potential to be used to support EFL learners of academic writing.

Keywords: *ChatGPT-assisted feedback; lexical resource; EFL argumentative writing; technology acceptance model; learning motivation; PLS-SEM.*

Introduction

The integration of AI in the classroom has influenced how feedback is achieved in EFL Writing instruction. AI-based tools help students receive more immediate feedback and have more autonomy and personalization in their learning. Amongst these tools, OpenAI's ChatGPT is regarded as a thoughtful technological advance. ChatGPT provides feedback and suggestions on incorrect answers or those that lack adequate context. AI-based classrooms are more engaging, largely due to the nature of the technology used (Loncar et al., 2023) and students' ability to learn at their own pace (Peterson, 2017). The combination of machine learning and natural language processing (NLP) creates more engaging ways for students to practice and develop language skills. It aids learner autonomy by providing writing-style revision

CITATION | Thang, H. M. (2026). ChatGPT-Assisted Feedback and the Development of Lexical Resources in EFL Argumentative Writing: Students' Perceptions and Motivation. Computer-Assisted Language Learning Electronic Journal (CALL-EJ), 27(3), 61-90. DOI

support (Guo et al., 2022). Writing repetition activities used as feedback in EFL writing have been aided by the introduction of ChatGPT. ChatGPT can provide grammar feedback and lexical suggestions based on vocabulary and word choice, making it useful for writing essays with argumentative/persuasive language. Many learners of English as a Foreign Language (EFL) still struggle with insufficient lexical resources. These challenges may be compounded by limitations in the timeliness and specificity of formative feedback, as Nguyen, T. H. O. (2024) found that some Vietnamese university students received delayed or insufficiently detailed feedback, which constrained their opportunities for effective revision of their writing. Although teachers' written feedback plays an important role in improving students' writing skills (Nguyen, T. L., 2024), students' experiences with feedback may vary depending on its timeliness, clarity, and the provision of constructive suggestions (Nguyen, T. L., 2024; Nguyen, T. H. O., 2024). Thus, there are clear limitations on students' opportunities to make the many revisions required to improve the quality of their writing, especially in vocabulary. One answer to this problem could be to incorporate ChatGPT feedback into the writing process as a replacement for, or a supplement to, the teacher's role. No matter how many times students engage with ChatGPT, they will still be encouraged to work more independently to assess and improve the lexical choices they have made in their writing. Of course, there are major safety concerns when considering the educational use of ChatGPT; for example, the loss of students' motivation to complete writing tasks independently, potential breaches of academic integrity, and the responsibilities of teachers, who must adjust their practice (Shoufan, 2023). These results show that the effect of ChatGPT in class is as important as the system itself. Moreover, most studies on ChatGPT focus either on students' perceptions or on learning outcomes. There have been few attempts to consider them jointly. Lastly, although students' acceptance of AI technologies has been studied, there is little empirical research showing how positive opinions towards ChatGPT have led to constructive changes in students' writing. Answering this question is essential because simply adding technology to a learning environment is unlikely to improve learning unless students actively engage with the technology and AI-based feedback.

This study aims to address these gaps by examining the use of ChatGPT as a pedagogically scaffolded feedback tool to stimulate the use of lexical resources in EFL argumentative essay writing. Using the Technology Acceptance Model (Davis, 1989), this study aims to determine whether participants in the ChatGPT feedback group exhibit greater lexical growth than those in the traditional teacher feedback group. Further, it would examine participants' opinions about ChatGPT regarding usefulness, ease of use, and learning motivation, and assess whether motivation to learn mediates the relationships between perceived usefulness of ChatGPT and lexical resources developed, and between perceived ease of use of ChatGPT and lexical resources developed. Instead of treating ChatGPT as a content-generation program, this study considers it a feedback tool that can assist students in revising their work while retaining students' accountability for planning, drafting, and revising their own work.

Accordingly, the study addresses the following research questions:

> RQ1. Do students who receive ChatGPT-assisted feedback demonstrate greater improvement in lexical resources in argumentative writing than students who receive traditional teacher feedback?
>
> RQ2. Do students in the ChatGPT-assisted feedback group report higher perceived usefulness, perceived ease of use, and motivation than students in the traditional feedback group?

RQ3. To what extent do perceived usefulness and perceived ease of use predict students' motivation to engage with ChatGPT-assisted feedback?

RQ4. Does motivation mediate the relationship between perceived usefulness, perceived ease of use, and lexical resource development?

Literature review

ChatGPT-Assisted Feedback in EFL Writing

Artificial intelligence (AI) has become increasingly prominent in language education, particularly in English as a Foreign Language (EFL) writing instruction. Among recent AI applications, ChatGPT has attracted increasing attention for its ability to generate context-sensitive responses and provide immediate feedback throughout the writing process. Unlike conventional automated writing evaluation systems that primarily identify grammatical errors, ChatGPT enables learners to interact with feedback through a conversational interface. Students can ask target questions, come up with different ways to say the same thing, and revise numerous drafts. Feedback becomes a more interactive and iterative process. According to recent studies, ChatGPT can aid in writing development by explaining feedback, selecting more appropriate terms, suggesting changes to sentence structure and phrasing, and suggesting revisions to the draft several times. (Han & Li, 2024; Polakova & Ivenz, 2024; Teng, 2024; Meniado et al., 2024). There are, however, many issues that are associated with the education sector as a whole, of which ChatGPT is a part. These include students over-relying on ChatGPT-suggested answers, a loss of integrity in completing academic tasks, and the absence of adequate educational resources to guide students (Lingard, 2023; Imran & Almusharraf, 2023). Consequently, many recent articles argue that, rather than replacing the task of independent writing, ChatGPT should assist by providing feedback. How much value it adds depends on its integration into classroom activities (Kasneci et al., 2023; Naweed, 2024). Besides providing linguistic assistance, ChatGPT may aid learners' active engagement in planning, monitoring, and revising their writing. From a sociocultural perspective, immediate and individualized AI responses may serve as instructional scaffolding that supports students within their Zone of Proximal Development (Huang A. Y. et al., 2023; Huang X. et al., 2023; Fulton et al., 2021). Although these perspectives offer useful teaching suggestions for using AI, the present study does not analyze self-regulated learning or the sociocultural perspective. Rather, these teaching suggestions are used to complement other AI-related suggestions, and the Technology Acceptance Model is the primary framework for explaining learners' acceptance of ChatGPT-assisted feedback.

Lexical Resource Development in Argumentative Writing

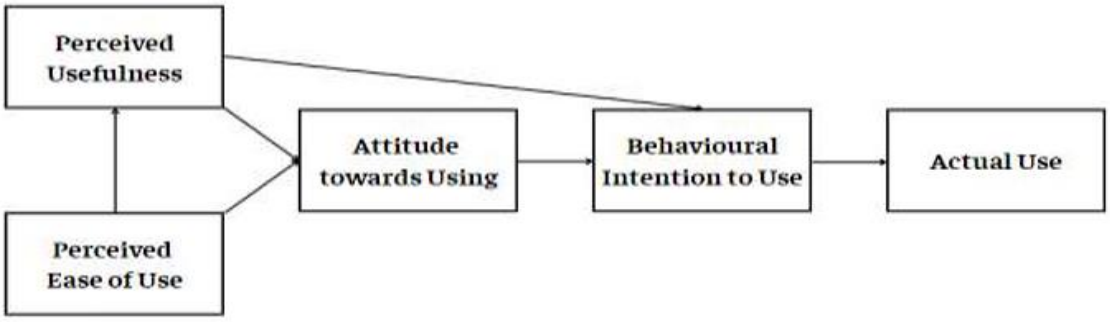
Lexical resources in writing mean using an appropriate variety of words and expressing ideas effectively in a particular context. In argumentative writing, lexical resources aid the presentation of ideas in a logical, easily interpretable way and persuade readers of the writer's point of view. A sufficient variety of lexical resources aids argument development and avoids distortion due to repetition (Muñoz, 2014; Cahyono & Widiati, 2009; Asyiah, 2017). Developing sufficient lexical resources for academic writing is problematic for most English as a Foreign Language (EFL) learners. Roughly speaking, these challenges become more evident in higher education settings due to limited opportunities for detailed lexical feedback. Recent studies have shown that the integration of AI provides learners with resources to develop vocabulary and assist in using more appropriate lexical items, offering definitions of lexical differences, and suggesting lexical

alternatives by means of contextual embedding (Kim, 2018; Lei, 2022; Oktadela et al., 2023; Abdelhalim & Alsehibany, 2025; Jmaiel et al., 2025). The general tendency in most previous studies is that AI-based feedback systems may assist in developing lexical resources for argumentative writing. However, this tendency remains unsubstantiated due to the lack of related studies.

Technology Acceptance Model

The Technology Acceptance Model (TAM) (Davis, 1989) says that perceived usefulness (PU) and perceived ease of use (PEOU) combine to determine users' technology acceptance. Both of these elements influence learners' Technology Acceptance Model (Venkatesh & Davis, 2000). Since ChatGPT has been confined to the learning technology domain, the likelihood of the system's acceptance will determine how students interact with AI-based revision-writing feedback. Learners tend to adopt technologies they perceive to be useful and easy to use. Positive perceptions of educational technologies lead to higher engagement, greater motivation, and sustained use of tools in educational environments. Positive perceptions of ChatGPT will lead learners to revise their writing more often, improving their word choice and expression. Also, perceived usefulness, ease of use, and writing quality are likely to play a significant role as mediating mechanisms.

Figure 1.
Technology Acceptance Model (Davis, 1989).



Research Gap and Hypotheses

Though research on ChatGPT-supported writing is growing, some gaps remain. Some of these gaps relate to the specific lexical resource and the research methods. For example, when comparing one of ChatGPT's advantages with another, previous research has focused more on the overall quality of the text produced. At the same time, some studies have analyzed the components of writing separately. Although the Technology Acceptance Model has frequently been applied to explain learners' acceptance of educational technologies, limited empirical evidence exists regarding the motivational mechanisms by which perceived usefulness and perceived ease of use contribute to lexical resource development in EFL writing. To address these gaps, the present study investigates the effects of ChatGPT-assisted feedback on lexical resource development. It examines whether learning motivation mediates the relationship between perceived usefulness, perceived ease of use, and lexical resource development. Based on the Technology Acceptance Model and previous empirical findings, the following hypotheses are proposed:

H1. Perceived usefulness positively influences students' motivation to use ChatGPT-assisted feedback.

H2. Perceived ease of use positively influences students' motivation.

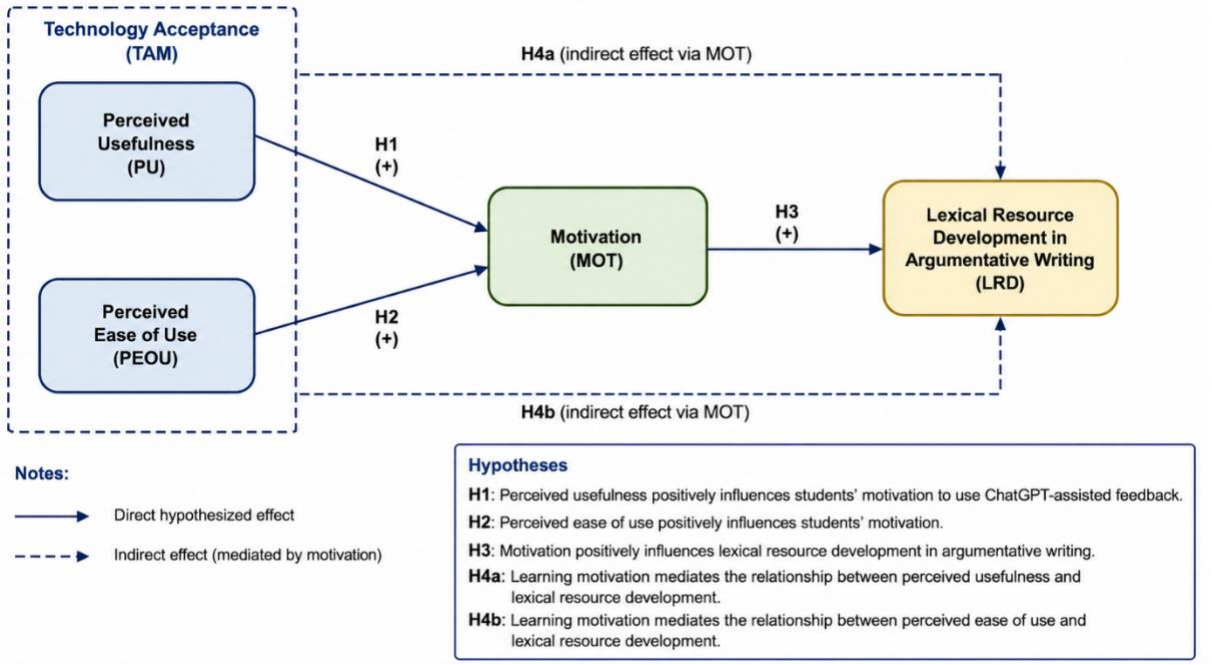
H3. Motivational factors positively influence lexical resource development.

H4a: Learning motivation mediates the relationship between perceived usefulness and lexical resource development.

H4b: Learning motivation mediates the relationship between perceived ease of use and lexical resource development

Figure 2.

Conceptual framework of the study.



Methods

Research Design

This study employed a quasi-experimental pre-test-post-test control-group design to investigate the effects of ChatGPT-assisted feedback on Vietnamese EFL students' lexical resource development in argumentative writing. Two intact undergraduate writing classes were assigned to either an experimental group receiving ChatGPT-assisted feedback or a control group receiving conventional teacher feedback. Because participants were assigned by class rather than through individual randomization, the study adopted a quasi-experimental design. The study was guided by the Technology Acceptance Model (Davis, 1989), which proposes that perceived usefulness and perceived ease of use influence learners' acceptance of educational technologies. Based on this framework, learning motivation was hypothesized to mediate the relationship between perceived usefulness, perceived ease of use, and lexical resource development.

Participants and Sampling Method

This study included 120 participants, who were Vietnamese undergraduates and took a writing course at the University of Economics Ho Chi Minh City (UEH). Participants had all completed the university's foundation English course and had an intermediate level of English (CEFR B1-B2). In this study, I used a data-gathering method that allows a researcher to recruit volunteers

without a particular selection process, called convenience sampling. I recorded two courses and put 60 participants in one course, which was the experimental group. Participants in this course received ChatGPT-assisted feedback, and 60 were assigned to the control group that received conventional feedback. Both groups went through the same processes and procedures for teaching and learning activities, the course syllabus, learning objectives, instructional schedule, writing tasks, and assessment criteria. The characteristics of the participants are in Table 1.

Table 1.

Participant Characteristics

Variable	Category	Experimental (n = 60)	Control (n = 60)	Total (N = 120)
Gender	Male	24 (40.0%)	26 (43.3%)	50 (41.7%)
	Female	36 (60.0%)	34 (56.7%)	70 (58.3%)
Age (years)	Mean (SD)	20.31 (0.82)	20.18 (0.79)	20.25 (0.81)
	Range	19-22	19-22	19-22
English proficiency (CEFR)	B1	35 (58.3%)	37 (61.7%)	72 (60.0%)
	B2	25 (41.7%)	23 (38.3%)	48 (40.0%)
Years of learning English	Mean (SD)	9.42 (2.01)	9.27 (1.95)	9.35 (1.98)
Previous experience using ChatGPT for learning	Yes	45 (75.0%)	47 (78.3%)	92 (76.7%)
	No	15 (25.0%)	13 (21.7%)	28 (23.3%)

Table 1 indicates that both groups have the same gender and age balance, as well as the same level of English proficiency and years of English learning. Also, about 77% of participants have used ChatGPT for their studies and general learning, but they did not use it to prepare academic texts or to receive feedback from AI. Providing AI-assisted feedback and using ChatGPT in learning activities gave the participants a similar background, helping minimize the effects of pre-study differences.

Research Procedure

The intervention took place over eight weeks during a single academic semester. Details are shown in Table 2. Both groups engaged in the same instructional activities, writing tasks, and cycles of revision. The only discrepancy between the two groups was the source of feedback for essay revision. At the beginning of the study, participants completed a pre-test argumentative writing task under the same testing conditions. During the intervention, students completed three argumentative writing tasks. Each of these tasks contained an initial draft and submission, along with feedback and a revision. After the intervention, all participants completed a post-test task, a writing prompt, and a survey to capture perceived ease of use, perceived usefulness, and motivation for the learning process. To maintain formality in the experimental group, students conversed with ChatGPT with a standard prompt protocol (Appendix B). Students in the control group received equivalent written lexical feedback from the course instructor.

Table 2.

Eight-week research procedure

Week	Experimental Group (ChatGPT-assisted Feedback)	Control Group (Teacher Feedback)
Week 1	Orientation to the study, informed consent, introduction to the research protocol and responsible use of ChatGPT as a lexical feedback tool. Completion of the pre-test argumentative writing task under examination conditions.	Orientation to the study, informed consent, introduction to the research procedures. Completion of the pre-test argumentative writing task under identical examination conditions.
Week 2	Instruction on argumentative writing and academic vocabulary. Students completed **Draft 1** independently before submitting it to ChatGPT using the standardized prompt. ChatGPT identified inappropriate or repetitive vocabulary, suggested more formal academic expressions, recommended natural collocations, and explained each suggestion (e.g., replacing *bad effects* with *negative consequences*). **Example:** Student text: *"Social media has many bad effects on students."* ChatGPT feedback: • Replace *bad effects* with *negative consequences* • Replace *many* with *numerous*	Instruction on argumentative writing and academic vocabulary. Students completed **Draft 1** independently and received written lexical feedback from the instructor. Teacher feedback focused on inappropriate vocabulary, academic word choice, collocations, and lexical accuracy (e.g., recommending the replacement of *bad effects* with *negative consequences*). **Example teacher comment:** *Consider replacing "bad effects" with a more academic expression such as "negative consequences." Avoid overusing common adjectives like "many" and use more precise vocabulary where appropriate.*
Week 3	Students critically evaluated ChatGPT's suggestions before revising **Draft 1**. They were encouraged to accept, modify, or reject each recommendation rather than applying all suggestions automatically. Revised drafts were submitted for assessment. **Example revision** Original: *Social media has many bad effects on students.* Revised: *Social media may have numerous negative consequences for university students,*	Students revised **Draft 1** independently after considering the teacher's written comments. The revised drafts reflected improvements in vocabulary choice, lexical precision, and academic style before submission. **Example revision** Original: *Social media has many bad effects on students.* Teacher feedback: *Replace "bad effects" with a more formal expression.* Revised:

	including reduced concentration and increased dependence on digital platforms.	*Social media has several negative consequences for university students.*
Week 4	Students completed **Draft 2** on a new argumentative topic. ChatGPT provided lexical feedback focusing on vocabulary choice, academic vocabulary, collocations, lexical variation, and paraphrasing while preserving the original meaning.	Students completed **Draft 2** on the same topic. The instructor provided equivalent written lexical feedback targeting vocabulary range, collocations, lexical appropriateness, and paraphrasing.
Week 5	Students revised **Draft 2** based on ChatGPT feedback, with particular attention to improving lexical sophistication, avoiding repetition, and selecting contextually appropriate academic vocabulary.	Students revised **Draft 2** using teacher feedback, focusing on improving vocabulary accuracy, lexical variation, and appropriate academic expressions.
Week 6	Students completed **Draft 3** on a different argumentative topic. ChatGPT again provided lexical feedback to help refine vocabulary choice, collocations, paraphrasing, and academic word use before revision.	Students completed **Draft 3** on the same topic and received teacher feedback addressing the same lexical dimensions to ensure comparability between the two groups.
Week 7	Students completed the final revision of **Draft 3**, reflecting on vocabulary improvements across the three writing cycles. The instructor monitored students' use of ChatGPT to ensure compliance with the research protocol.	Students completed the final revision of **Draft 3** following teacher feedback and reflected on their vocabulary development throughout the writing tasks.
Week 8	Students completed the post-test argumentative writing task under examination conditions without using ChatGPT or other AI tools. They then completed the Technology Acceptance and Learning Motivation Questionnaire.	Students completed the post-test argumentative writing task under the same examination conditions, followed by completion of the same questionnaire.

Note. Both groups followed the same syllabus, writing topics, instructional schedule, number of drafts (three), revision cycles (three), and assessment criteria. The experimental group received feedback from ChatGPT, while the control group received feedback from teachers. Examples of the standardized ChatGPT prompt, ChatGPT feedback, teacher feedback, and student revisions are available in Appendix B.

Intervention and Feedback Protocols

Experimental Group: ChatGPT-Assisted Feedback

All students attended a training session before the intervention. They were instructed to use ChatGPT only for lexical feedback. Students were instructed to draft independently and then submit the draft for feedback using a standardized feedback protocol. ChatGPT was asked to suggest vocabulary and paraphrases, and to provide alternative phrasings and expressions, without the ability to generate suggested responses or rewrite the draft or essay. ChatGPT suggested

revising "bad effects" to "negative consequences" and again suggested using "numerous" instead of "many". ChatGPT justified replacing "bad effects" and "many" with "negative consequences" and "numerous". Students critically assessed each suggested change and made the final determination on whether to adopt it. Students did not unquestioningly accept suggestions; rather, they accepted, revised, or rejected the suggestion based on their WRITS. The instructor tracked the use of ChatGPT to ensure compliance with the protocol, and Appendix B contains the standardized prompt, usage instructions, examples of feedback, and student revisions.

Control Group: Teacher Feedback

Students in the control group were given the same writing tasks and schedules as the students in the treatment group. Rather than receiving feedback from ChatGPT, the control group students received written lexical feedback from the course instructor. The goal of this feedback was to highlight vocabulary, correct collocations, and use more formal language and paraphrasing. After receiving feedback, students revised their essays independently and submitted them to the course instructor for a final grade.

Table 3.
Comparison of feedback conditions

Aspect	Experimental Group	Control Group
Feedback provider	ChatGPT	Course instructor
Writing tasks	Same	Same
Writing topics	Same	Same
Number of drafts	3	3
Number of revisions	3	3
Feedback focus	Lexical resource	Lexical resource
Vocabulary suggestions	✓	✓
Academic word choice	✓	✓
Collocation suggestions	✓	✓
Paraphrasing suggestions	✓	✓
Grammar correction	Limited	Limited
Essay generation	Not permitted	Not applicable
Example feedback	AI suggested replacing "*bad effects*" with "*negative consequences*" and explained the lexical improvement.	Teacher highlighted inappropriate vocabulary and recommended replacing "*bad effects*" with "*negative consequences*".

Research Instruments

Writing Test

Lexical resource performance was assessed using a study-adapted analytic rubric informed by the IELTS Writing Task 2 Lexical Resource band descriptors. Four dimensions were evaluated: Vocabulary Range, Word Choice Accuracy, Academic Vocabulary and Collocations, and Lexical Flexibility. Each dimension was rated on a nine-band scale (1–9), with Band 0 reserved for responses containing no assessable language. The overall lexical resource score for each essay was

calculated as the arithmetic mean of the four dimension scores, resulting in a composite score ranging from 1.00 to 9.00 for assessable responses. Two experienced EFL writing instructors independently evaluated the essays following a calibration session. Differences in ratings were resolved through discussion, and the agreed dimension scores were used to calculate the final composite score. This adapted analytic scoring procedure was developed for the purposes of the study and should not be interpreted as the official IELTS Writing Task 2 scoring procedure. The complete rubric and illustrative scoring examples are provided in Appendix D.

Questionnaire

After the intervention, students filled out our survey to share their thoughts on the ChatGPT-generated feedback. The survey was created based on the Technology Acceptance Model (Davis, 1989) and previous work on other AI-based tools for language learning (Losi et al., 2024). The survey consisted of three latent constructs: Perceived Usefulness (PU), Perceived Ease of Use (PEOU), and Learning Motivation (MOT). All items used a five-point Likert scale ranging from 1 (strongly disagree) to 5 (strongly agree). Perceived Usefulness measured students' beliefs on the role that ChatGPT-generated feedback can play in assisting students in developing their academic vocabulary and writing. Perceived Ease of Use examined how students perceived the simplicity and usability of ChatGPT in the writing revision process. Learning Motivation assessed students' willingness to use ChatGPT-generated feedback, their focus during the revision process, and their belief in the tool's potential to improve their lexical skills. A detailed survey with all the items is in Appendix A.

Validity and Reliability

Before analyzing the structural relationships, the measurement model was assessed to determine the instrument's validity and reliability. For both, the minimum was set to .70. Cronbach's Alpha and Composite Reliability (CR) were used to assess internal consistency, and values of both were assessed to see if they were equal to or greater than .70. Convergent validity was assessed using the factors' loading and Average Variance Extracted (AVE). Per the guiding principles, if the factor loadings were greater than .70 and the AVE was greater than .50, they were considered adequate. Discriminant validity was assessed through the Heterotrait Monotrait (HTMT) ratio. After all these assessments, each construct was evaluated to ensure it measured one and only one construct, and only then were the structural relationships among the latent variables assessed.

Data Analysis

Quantitative data analysis was conducted using SPSS Statistics 26 and SmartPLS 4. SPSS was used for preliminary statistical analysis. This included descriptive statistics, independent-samples t-tests, and paired-samples t-tests. SmartPLS 4 was used to evaluate the proposed relationships among the latent variables using PLS-SEM. The analysis was conducted in four steps. Data visualizations were performed using SPSS Statistics 26 and Microsoft Excel. Following the data analyses, charts and tables were generated to visualize the results. The final step was to summarize the findings from the collected data. Each hypothesis was assessed to determine whether the data sufficiently supported it. If unsupported, the context in which the hypothesis was formed was presented.

First, descriptive statistics were calculated to summarize participants' demographics and survey responses. Mean and standard deviation, frequency, and percentage calculations were made. Before considering the effects of the intervention, an independent-samples t-test was performed to

determine whether the two groups (i.e., experimental and control) were equivalent at the pre-test stage. Next, changes in lexical resource scores were examined using paired-samples t-tests within each group, and independent-samples t-tests were conducted to compare post-test scores and learning gains between the two groups. In addition, Cohen's d was calculated to estimate the intervention's effect size. After that, the measurement model was evaluated for reliability and validity, with the structural relationships examined. Internal consistency was measured using Cronbach's alpha and Composite Reliability (CR), with values greater than .70 accepted for both. Convergent validity was also evaluated by standardized factor loadings and Average Variance Extracted (AVE). According to Hair et al. (2022), factor loadings above .70 and AVE values above .50 confirm convergent validity. Discriminant validity was determined by the Heterotrait-Monotrait (HTMT) ratio to ensure that each construct measured a unique theoretical concept. Lastly, the structural model was evaluated to examine the relationships between perceived usefulness, perceived ease of use, motivation, and lexical resource development. Before hypothesis testing, the potential of each predictor construct was assessed using Variance Inflation Factors (VIFs). The model's strength and explanatory power were also assessed using the coefficient of determination (R^2), and predictive relevance was assessed using Stone-Geisser's Q^2 values. Standardized path coefficients (β), t-values, and p-values were computed to analyze the relationships specified in hypotheses 1 to 3. The last step was to examine the role of motivation using a bootstrapping method with 5,000 resamples. Both direct and indirect effects were analyzed, and if the 95% bias-corrected confidence interval of the indirect effect did not include zero, mediation was considered significant. This method enabled the research to determine whether motivation mediated the relationships among perceived usefulness, perceived ease of use, and lexical resource development. Model fit was assessed using the Standardized Root Mean Square Residual (SRMR) and the Normed Fit Index (NFI), as suggested in the PLS-SEM literature.

Lexical resource performance was operationalized as a composite score derived from four study-specific analytic dimensions, each rated on a nine-band scale informed by the IELTS Writing Task 2 Lexical Resource descriptors. The composite score was calculated by averaging the four agreed dimension scores. Individual learning gains were calculated by subtracting pre-test composite scores from post-test composite scores. These composite scores, rather than official IELTS Writing Task 2 band scores, were used in the statistical analyses.

Ethical Considerations

Informed consent was given before data collection. Participants were told about the study and assured that the data would be kept confidential and no names would be recorded. Participants were told that they could leave the study at any time without punishment. ChatGPT was used strictly for research and writing feedback, in accordance with the instructions and the university's policy on responsible AI use.

Findings

Preliminary Analysis

A total of 120 students participated in the study and completed the pre-test, post-test, intervention, and questionnaire. A pre-intervention analysis was conducted to determine whether the participants in the experimental and control groups were comparable. An independent-samples t-test found no significant difference in lexical resource scores on the pre-test between the two groups ($p > .05$), indicating that participants in the control and experimental groups had the same

level of writing proficiency before the intervention. This provided a basis for the analysis of the post-intervention lexical resource scores.

Effects of ChatGPT-Assisted Feedback on Lexical Resource Development

Table 4.

Pre-test and Post-test Lexical Resource Scores of the Experimental and Control Groups

Group	n	Pre-test Mean (SD)	Post-test Mean (SD)	Mean Gain	t-value	p-value	Cohen's *d*
Experimental (ChatGPT-assisted feedback)	60	5.92 (0.79)	7.74 (0.68)	+1.82	14.36	< .001	1.85
Control (Teacher feedback)	60	5.94 (0.83)	6.38 (0.76)	+0.44	3.21	< .01	0.42

Note. M = mean; SD = standard deviation. Lexical resource scores were calculated using a study-adapted analytic rubric informed by the IELTS Writing Task 2 Lexical Resource band descriptors. Each of the four dimensions was rated on a nine-band scale, and the overall score was calculated as the arithmetic mean of the four dimension scores. Mean gain represents the difference between post-test and pre-test composite scores. The reported scores are research-specific composite scores rather than official IELTS Writing Task 2 band scores.

As shown in Table 4, both groups demonstrated improvements in lexical resource scores from pre-test to post-test. The experimental group, which received ChatGPT-assisted feedback, increased from M = 5.92 (SD = 0.79) to M = 7.74 (SD = 0.68), representing a mean gain of 1.82 points, $t(59) = 14.36$, $p < .001$. The control group, which received teacher feedback, improved from M = 5.94 (SD = 0.83) to M = 6.38 (SD = 0.76), with a smaller mean gain of 0.44 points, $t(59) = 3.21$, $p < .01$. Descriptively, the experimental group's mean gain exceeded that of the control group by 1.38 points. However, a between-group comparison of individual gain scores is necessary to establish whether this difference is statistically significant.

Technology Acceptance and Learning Motivation

Table 5.

Students' Perceptions of ChatGPT-Assisted Feedback in the Experimental Group

Construct	Group	N	Mean (M)	Standard Deviation (SD)	t-value	p-value
Perceived Usefulness (PU)	Experimental	60	4.52	0.41	9.04	< .001
Perceived Ease of Use (PEOU)	Experimental	60	4.37	0.40	6.10	< .001

Learning Motivation	Experimental	60	4.61	0.43	9.82	< .001

Note. PU = Perceived Usefulness; PEOU = Perceived Ease of Use. Independent-samples t-tests were conducted to compare post-test perception scores between the experimental and control groups. All items were measured on a five-point Likert scale (1 = strongly disagree, 5 = strongly agree).

Table 5 presents students' post-intervention opinions on the use of ChatGPT-assisted feedback. Students in the experimental group expressed more positive opinions than those in the control group regarding perceived usefulness, instrument ease of use, and learning motivation (all $p < .001$). Among these, learning motivation with a mean score of $M = 4.61$ ($SD = 0.43$) was the highest, followed by perceived usefulness, $M = 4.52$ ($SD = 0.41$), and perceived ease of use, $M = 4.37$ ($SD = 0.40$). Generally, students expressed positive views on the use of ChatGPT-assisted feedback in argumentative writing.

Measurement Model Assessment

After analyzing the perceptions of the surveyed students, the measurement model was assessed before evaluating the structural relationships proposed in the conceptual framework. Measurement model assessment is done in the first stage of the two-stage approach to Partial Least Squares Structural Equation Modeling (PLS-SEM). Following the recommendation, the first assessment of the measurement model is to evaluate the reliability and validity of latent constructs. Reliability, convergent validity, and construct validity are assessed by Cronbach's alpha, Composite Reliability (CR), Average Variance Extracted (AVE), and standardized factor loadings, respectively. The results of the assessment of the measurement model are presented in Table 6.

Table 6.

Measurement Model Assessment

Construct	No. of Items	Factor Loading Range	Cronbach's Alpha	Composite Reliability (CR)	AVE
Perceived Usefulness (PU)	4	0.78-0.91	0.89	0.92	0.74
Perceived Ease of Use (PEOU)	4	0.75-0.88	0.87	0.91	0.71
Learning Motivation	5	0.79-0.92	0.91	0.93	0.76
Lexical Resource Development	4	0.77-0.89	0.88	0.91	0.73

Note. All standardized factor loadings exceeded .70. Cronbach's alpha and Composite Reliability (CR) values were above the recommended threshold of .70. In contrast, Average Variance Extracted (AVE) values exceeded .50, indicating satisfactory reliability and convergent validity.

As presented in Table 6, all constructs demonstrated satisfactory internal consistency. Cronbach's alpha values ranged from 0.87 to 0.91, while Composite Reliability (CR) values ranged from 0.91 to 0.93, both exceeding the recommended threshold of 0.70 (Hair et al., 2022). Convergent validity was also established, as all standardized factor loadings exceeded 0.70 and the Average Variance

Extracted (AVE) values ranged from 0.71 to 0.76, both exceeding the recommended minimum of 0.50 (Hair et al., 2022).

Following confirmation of the measurement model's reliability and validity, the proposed structural relationships were examined using a bootstrapping procedure with 5,000 resamples. After establishing internal consistency and convergent validity, discriminant validity was assessed using the Heterotrait–Monotrait ratio (HTMT). The HTMT values are presented in Table 7.

Table 7.

Discriminant Validity (HTMT)

Construct	PU	PEOU	Motivation	Lexical Resource Development
PU	—			
PEOU	0.62	—		
Motivation	0.71	0.65	—	
Lexical Resource Development	0.58	0.54	0.73	—

Note. All HTMT values were below the recommended threshold of .85, confirming satisfactory discriminant validity.

As shown in Table 7, all HTMT values ranged from 0.54 to 0.73, remaining below the recommended threshold of 0.85 (Henseler et al., 2015; Hair et al., 2022). These findings confirm satisfactory discriminant validity, indicating that each construct measured a conceptually distinct latent variable. Collectively, the measurement model demonstrated satisfactory psychometric properties and was therefore suitable for subsequent structural model analysis.

Structural Model and Hypothesis Testing

Following confirmation of the measurement model, the structural model was evaluated using the bootstrapping procedure with 5,000 resamples. Before hypothesis testing, collinearity diagnostics were examined to ensure that multicollinearity was not a concern. All Variance Inflation Factor (VIF) values were below 3.00, satisfying the recommended threshold for PLS-SEM analyses (Hair et al., 2022). Consequently, no multicollinearity issues were detected among the predictor constructs. The structural model demonstrated satisfactory explanatory power, accounting for 61% of the variance in lexical resource development ($R^2 = .61$). According to Hair et al. (2022), an R^2 value of 0.61 indicates moderate-to-substantial explanatory power. Model fit was also acceptable, with a Standardized Root Mean Square Residual (SRMR) of 0.061, which was below the recommended threshold of 0.08, and a Normed Fit Index (NFI) of 0.91, exceeding the recommended benchmark of 0.90 (Hair et al., 2022). The results of the structural model and hypothesis testing are shown in Table 8. According to Table 8, learning motivation was positively and significantly driven by perceived usefulness ($\beta = 0.49$, $t = 8.12$, $p < .001$) and perceived ease of use ($\beta = 0.33$, $t = 6.44$, $p < .001$). Consequently, Hypotheses 1 and 2 were supported. Learning motivation was positively and significantly related to the development of lexical resources ($\beta = 0.56$, $t = 9.27$, $p < .001$), supporting Hypothesis 3.

The effects of learning motivation on the perceived usefulness and perceived ease of use were examined by bootstrapping with 5,000 resamples. It was found that perceived usefulness exerted

a significant indirect effect on the development of lexical resources (indirect β = 0.274, 95% CI [0.16, 0.43]), and learning motivation partially mediated this effect. Also, perceived ease of use exerted a significant indirect effect on the development of lexical resources (indirect β = 0.185, 95% CI [0.08, 0.30]), and learning motivation partially mediated this effect. The direct effects of perceived usefulness and perceived ease of use on the development of lexical resources were β = 0.18 (t = 2.67, p = .008) and β = 0.13 (t = 2.18, p = .029), respectively, supporting Hypotheses 4a and 4b.

Table 8.

Results of Structural Model and Hypothesis Testing

Hypothesis	Path	β	t	p	Indirect β	95% CI	Decision
H1	PU → Motivation	0.49	8.12	<.001	—	—	Supported
H2	PEOU → Motivation	0.33	6.44	<.001	—	—	Supported
H3	Motivation → Lexical Resource Development	0.56	9.27	<.001	—	—	Supported
H4a	PU → Motivation → Lexical Resource Development	0.18	2.67	.008	0.274	[0.16, 0.43]	Supported (Partial mediation)
H4b	PEOU → Motivation → Lexical Resource Development	0.13	2.18	.029	0.185	[0.08, 0.30]	Supported (Partial mediation)

Note. Learning motivation mediated the relationships between perceived usefulness and lexical resource development, and between perceived ease of use and lexical resource development.

Model Quality Indicators

Indicator	Value
R^2	0.61
SRMR	0.061
NFI	0.91
VIF	< 3.00

Discussion

ChatGPT-assisted Feedback Improves Lexical Resource Development

The lexical gain shown by the experimental group indicates that ChatGPT's feedback goes beyond simply suggesting corrections. Compared with replacing teacher feedback, ChatGPT aimed to create opportunities for students to notice lexical problems, choose alternative ways to express

their ideas, and draft several versions of the text. This iterative behavior is certainly beneficial for students and would lead to deeper lexical processing, a larger lexicon, and a greater range of choices when writing. This process helps students develop writing skills and refine their use of lexicon. This means that vocabulary revision with the aid of AI, in this case ChatGPT in particular, was extremely useful for the participants. This illustrates that participants were using multiple drafts of their writing and revising lexical choices, coming up with alternatives and varying their lexicon across drafts. It also resonates with other studies, which show that writing quality is improved by using ChatGPT and other AI-assisted feedback, as it provides language scaffolding and encourages editing and revising the text several times (Han and Li, 2024; Polakova and Ivenz, 2024; Teng, 2024; Meniado et al, 2024). It also shows a pattern of increased lexical engagement, as seen in AI-generated feedback reported by Rahman et al. (2022) and Fitria (2021), in which students began revising their work several times to test the boundaries of vocabulary and sentence construction before finalizing the draft. This study is among the first to examine lexicon and vocabulary choice in argumentative writing, extending prior studies that did not focus on this aspect.

The results show that ChatGPT helps learners select appropriate words, borderline collocations, and academic phrases while preserving meaning, with little to no change in the original phrasing. This process has helped learners understand the importance of vocabulary choice. Beyond the correct lexicon in this context, ChatGPT's incremental feedback model may have created an environment in which students could process the lexicon more deeply. Students had to evaluate many AI feedback suggestions, compare lexicon alternatives, contemplate lexicon appropriateness, revise independently, and make their own decisions several times during the process. This continuous structural lexical feedback in writing tasks may have enhanced lexical awareness and contributed to more sustained lexical growth during subsequent writing tasks.

Technology acceptance and learning motivation

The second important finding focuses on students' acceptance of ChatGPT-generated feedback. Following the intervention, participants reported high levels of perceived usefulness and perceived ease of use. Both constructs were found to predict learning motivation. The finding confirms the Technology Acceptance Model (Davis, 1989), which posits that learners accept educational technologies they perceive as useful and easy to use. The findings are also consistent with prior literature documenting that positive perceptions of AI-integrated learning frameworks enhance learners' disposition to engage in writing tasks and sustain their effort to learn (Kasneci et al., 2023; Naweed, 2024). Furthermore, the immediate nature of the feedback provided by ChatGPT may explain students' perception of the system as both useful and easy to use. At the same time, they engaged in writing. Although perceived usefulness and perceived ease of use are likely to be considered in the context of technology adoption, the findings of the current study suggest that the two constructs may be instrumental in complementary and unique ways in motivating students. Specifically, students who perceived ChatGPT as useful motivated themselves to revise their vocabulary in the expectation of substantial learning, while students who perceived ChatGPT as convenient motivated themselves to engage with feedback at every step of the writing process. The current study builds on prior research by suggesting that incorporating AI into the classroom is vital and that perceived ease of use and perceived usefulness should be emphasized. In this study, perceived usefulness was rated on average higher than perceived ease of use. This shows that students valued ChatGPT because of its ability to enhance their vocabulary and improve their writing. This shows that the technology had a positive effect on their studies beyond just being

convenient. This is consistent with previous studies done using the Technology Acceptance Model, which found that most students typically consider the positive effect technology will have on their studies to be greater than the ease of use.

The mediating role of learning motivation

The structural model also confirmed that learning motivation moderated the relationship between perceived usefulness and the development of lexical resources. Similarly, learning motivation moderated the relationship between perceived ease of use and lexical resource development. For students who perceived ChatGPT as useful and easy to use, motivation to learn was higher and led to greater lexical resource development during the revision process. This reveals that the positive impact of ChatGPT on education cannot be explained solely in terms of technology and tools. The motivational level of the student users plays a big role in the positive impact of these technologies. Prior studies have consistently shown positive attitudes toward AI-based technologies. However, defining the mechanisms that underpin positive attitudes, perceptions of usefulness, and ease of use, and their eventual impact on learning outcomes, remains the biggest gap in existing studies. The present study, by incorporating learning motivation into the well-known Technology Acceptance Model, explains how feedback from ChatGPT facilitates vocabulary development in EFL writing. This study indicates that motivation acts as the connection between educational technologies and students' outcomes. Finally, the findings of this study confirm that, within the Technology Acceptance Model, perceived usefulness and perceived ease of use of a technology affect learning outcomes. In contrast, the motivational processes mediate this impact, in contrast with other studies in which motivational processes are related to the use of educational technologies. As such, the findings demonstrate that the psychological mechanism by which students' positive perceptions of ChatGPT translate into measurable gains in lexical resource development is learning motivation. This finding also provides a more comprehensive explanation of how AI-assisted feedback contributes to writing performance in EFL contexts.

Implications for AI-assisted EFL writing

The outcomes suggest that ChatGPT is likely to be most valuable when integrated into a prescribed feedback system rather than for generating written content. In our study, we instructed students to use ChatGPT exclusively for lexical feedback while revising. Accordingly, students were responsible for planning, drafting, and revising their essays. The direction given to the students likely contributed to improvements in their use of lexical resources, without the students viewing ChatGPT as a substitute for independent writing. The findings also contribute to the existing literature on AI-mediated language learning and underscore the necessity for careful consideration of instructional design. While participants' perceptions of ChatGPT were generally positive, our study indicates that positive perceptions do not automatically improve learning. Rather, students' motivation is the primary factor determining the extent of improvement in their writing. This adds to the current discourse suggesting that graduate students should not view generative AI systems and similar programs as substitutes for their own work and cognitive effort in the learning process (Kasneci et al., 2023; Lingard, 2023; Naweed, 2024). Our study is preliminary. Therefore, the findings of our study must be interpreted within the study's limitations. The sentences, instructional activities, tasks, and other aspects of teaching were created by the researcher. The main goal of the research was to develop more effective methods for teaching lexical resources in argumentative writing in Vietnamese universities. Based on our study's findings, future research should focus on teaching students to critically evaluate lexical proposals from AI systems. Comparing different vocabulary options. Justifying changes made. And deciding why one piece of academic language

is better than another depending on the context. All of these can help develop vocabulary, metalinguistic awareness, and learners' autonomy. While ChatGPT can still correct writing and fill in some gaps, most studies, though not entirely positive, have discussed the impact AI assistance has had on writing. ChatGPT can be exhaustive for some learners and can encourage students to write less. Critical thinking and drafting with revisions can be negatively impacted. ChatGPT was used in this study as a lexical feedback tool. To create a first draft and make the final revisions, students had to do the rest. This allowed students to be actively engaged throughout the writing process. Further study can examine the frequency and duration of the intervention to see whether similar effects occur with other writing tasks, different teaching practices, or other educational activities.

Conclusion

This study evaluated how using ChatGPT for feedback on EFL argumentative writing affected the development of lexical resources, and how the variables of perceived usefulness, perceived ease of use, and learning motivation correlated with learning outcomes. With a quasi-experimental design coupled with PLS-SEM, the study examined two groups of students: one that received feedback via ChatGPT and the other that received traditional feedback from their teachers. To gauge the effectiveness of the eight-week intervention, the ChatGPT group was compared with the teacher-feedback group. The group that received feedback from ChatGPT developed lexical resources to a much greater extent than the teacher-feedback group. Using ChatGPT was perceived as highly useful and easy to use and resulted in high learning motivation. The final analysis using structural equation modeling revealed that learning motivation mediated the relationship between perceived usefulness and ease of use and lexical resource development. This revealed that a positive perception of ChatGPT had a direct effect on the development of lexical resources and an indirect effect through increased learning motivation. This study makes three contributions. The first is to extend research on lexical resource development rather than on writing evaluation as a whole. The second is adapting the Technology Acceptance Model alongside learning motivation. The final contribution is to use the first research from Vietnam on AI writing to provide empirical context in this area. While these advances have been made, there are limitations. This study was done at a single university using convenience sampling, which affects how the results can and cannot be applied. This study tested lexical resource development in argumentative writing for eight weeks. Future studies could build on existing research by focusing on longer intervention durations, involving participants from multiple educational contexts, and identifying the impacts of ChatGPT-assisted feedback on the organization and coherence of writing, as well as on grammar and critical thinking.

Acknowledgments

This research was funded by the University of Economics Ho Chi Minh City, Vietnam.

References

Abdelhalim, S. M., & Alsehibany, R. (2025). *Integrating ChatGPT for vocabulary learning and retention: A classroom-based study of Saudi EFL learners*. *Language Learning & Technology, 29*(1), 1-24. https://doi.org/10.64152/10125/73635

Asyiah, D. N. (2017). The vocabulary teaching and vocabulary learning: Perception, strategies, and influences on students' vocabulary mastery. *Jurnal Bahasa Lingua Scientia, 9*(2), 293-318. https://doi.org/10.21274/ls.2017.9.2.293-318

British Council, IDP: IELTS Australia, & Cambridge Assessment English. (2023). *IELTS Writing band descriptors (Task 2).* https://ielts.org

Cahyono, B. Y., & Widiati, U. (2009). The teaching of EFL vocabulary in the Indonesian context: The state of the art. *TEFLIN Journal, 19*(1), 1-17. https://doi.org/10.15639/teflinjournal.v19i1/1-17

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

Fitria, T. N. (2021). Grammarly as AI-powered English writing assistant: Students' alternative for writing English. *Metathesis: Journal of English Language, Literature, and Teaching, 5*(1), 65-78. https://doi.org/10.31002/metathesis.v5i1.3519

Fulton, L., Hoffman, D. L., & Paek, S. (2021). Language learning in an era of datafication and personalized learning. *Educational Technology Research and Development, 69*(1), 315-318. https://doi.org/10.1007/s11423-020-09903-0

Guo, K., Jian, W., & Chu, S. K. W. (2022). Using chatbots to scaffold EFL students' argumentative writing. *Assessing Writing, 54*, 100666. https://doi.org/10.1016/j.asw.2022.100666

Hair, J. F., Hult, G. T. M., Ringle, C. M., & Sarstedt, M. (2022). *A primer on partial least squares structural equation modeling (PLS-SEM)* (3rd ed.). Sage.

Han, J., & Li, M. (2024). Exploring ChatGPT-supported teacher feedback in the EFL context. *System, 126*, 103502. https://doi.org/10.1016/j.system.2024.103502

Huang A. Y., Lu O. H., Yang S. J. (2023). Effects of artificial intelligence-enabled personalized recommendations on learners' learning engagement, motivation, and outcomes in a flipped classroom.*Comput. Educ.*194:104684. https://doi.org/10.1016/j.compedu.2022.104684

Huang, X., Zou, D., Cheng, G., Chen, X., & Xie, H. (2023). Trends, Research Issues and Applications of Artificial Intelligence in Language Education. *Educational Technology & Society, 26*(1), 112-131. https://doi.org/10.30191/ETS.202301_26(1).0009

Imran, M., & Almusharraf, N. (2023). Analyzing the role of ChatGPT as a writing assistant at the higher education level: A systematic review of the literature. *Contemporary Educational Technology, 15*(4), ep464. https://doi.org/10.30935/cedtech/13605

Jmaiel, H. A., Abukhait, R. O., Mohamed, A. M., Shaaban, T. S., & Al-Qadri, A. (2025). The role of ChatGPT in enhancing EFL students' ESP writing skills: An experimental study of gender and major differences. *Discover Education*. https://doi.org/10.1007/s44217-025-00700-6

Kasneci, E., Küchemann, S., Bannert, M., Dementieva, D., Fischer, F., Gasser, U., Groh, G., Günnemann, S., & Hüllermeier, E. (2023). Opportunities and challenges of large language models for education. *OSF Preprints*. https://doi.org/10.35542/osf.io/5er8f

Kim, N. Y. (2018). Chatbots and Korean EFL students' English vocabulary learning. *Journal of Digital Convergence, 16*(2), 1-7. https://doi.org/10.14400/JDC.2018.16.2.001

Lei, H. (2022). High school students' foreign language vocabulary acquisition in the era of artificial intelligence. In *Proceedings of the International Conference on Education, Language and Art* (pp. 582-585). Atlantis Press.

Lingard, L. (2023). Writing with ChatGPT: An illustration of its capacity, limitations & implications for academic writers. *Perspectives on Medical Education, 12*(1), 261-270. https://doi.org/10.5334/pme.1072

Loncar, M., Schams, W., & Liang, J. S. (2023). Multiple technologies, multiple sources: Trends and analyses of the literature on technology-mediated feedback for L2 English writing.

Computer Assisted Language Learning, 36(4), 722-784. https://doi.org/10.1080/09588221.2021.1943452

Losi, R. V., Putra, E., Ali, N., & Dewi, A. S. (2024). Investigating Artificial Intelligence (AI) as a Vocabulary Learning Tool: Students' Perception to Use ChatGPT. *Proceeding International Conference on Religion, Science and Education*, *3*, 561–566. Retrieved from https://sunankalijaga.org/prosiding/index.php/icrse/article/view/1208

Meniado, J. C., Huyen, D. T. T., Panyadilokpong, N., & Lertkomolwit, P. (2024). Using ChatGPT for second language writing: Experiences and perceptions of EFL learners in Thailand and Vietnam. *Computers and Education: Artificial Intelligence, 7*, 100313. https://doi.org/10.1016/j.caeai.2024.100313

Muñoz, C. (2014). Exploring young learners' awareness of foreign language learning. *Language Awareness, 23*(1-2), 24-40. https://doi.org/10.1080/09658416.2013.863900

Naweed, S. (2024). Exploring teacher attitudes towards ChatGPT: A comprehensive review. *International Journal of Social Science and Entrepreneurship, 4*(1), 212-225. https://doi.org/10.58661/ijsse.v4i1.259

Nguyen, T. H. O. (2024). The impact of feedback on first-year university students' writing development: A study at a university in Hanoi, Vietnam. *International Journal of Social Science and Human Research, 7*(8). https://doi.org/10.47191/ijsshr/v7-i08-107

Nguyen, T. L. (2024). Analysis of teachers' written feedback on writing skills of third-year English-majored students at a Vietnamese university. *VNU Journal of Foreign Studies, 40*(1), 103-123. https://doi.org/10.63023/2525-2445/jfs.ulis.5188

Oktadela, R., Elida, Y., & Ismail, S. (2023). Improving English vocabulary through an artificial intelligence (AI) chatbot application. *Journal of English Language and Education, 8*(2), 63-67. https://doi.org/10.31004/jele.v8i2.308

Polakova, P., & Ivenz, P. (2024). The impact of ChatGPT feedback on the development of EFL students' writing skills. *Cogent Education, 11*(1). https://doi.org/10.1080/2331186X.2024.2410101

Rahman, N. A. A., Zulkornain, L. H., & Hamzah, N. H. (2022). Exploring artificial intelligence using automated writing evaluation for writing skills. *Environment and Behavior Proceedings Journal, 7*(SI9), 547-553. https://doi.org/10.21834/ebpj.v7iSI9.4304

Shoufan, A. (2023). Exploring students' perceptions of ChatGPT: Thematic analysis and follow-up survey. *IEEE Access, 11*, 38805-38818. https://doi.org/10.1109/ACCESS.2023.3268224

Teng, M. F. (2024). "ChatGPT is the companion, not enemies": EFL learners' perceptions and experiences in using ChatGPT for feedback in writing. *Computers and Education: Artificial Intelligence, 7*, 100270. https://doi.org/10.1016/j.caeai.2024.100270

Venkatesh, V., & Davis, F. D. (2000). A theoretical extension of the Technology Acceptance Model: Four longitudinal field studies. *Management Science, 46*(2), 186-204. https://doi.org/10.1287/mnsc.46.2.186.11926

Biodata

Thang Ho is an English lecturer at the University of Economics Ho Chi Minh City (UEH), Vietnam. With research interests in TELL, CALL, MALL, digital learning, self-regulated learning, and learner autonomy, he aims to inspire students through engaging, supportive classrooms that foster confidence, creativity, and cultural curiosity.

APPENDICES

Appendix A. Questionnaire Items

Perceived Usefulness

PU1. ChatGPT feedback helped me improve my academic vocabulary.
PU2. ChatGPT feedback improved the quality of my argumentative writing.
PU3. ChatGPT feedback supported my revision process.
PU4. Overall, ChatGPT was useful for improving my writing.

Perceived Ease of Use

PEOU1. Learning to use ChatGPT was easy.
PEOU2. Interacting with ChatGPT was clear and understandable.
PEOU3. ChatGPT was easy to use during writing revision.
PEOU4. I felt comfortable using ChatGPT independently.

Learning Motivation

MOT1. ChatGPT motivated me to revise my essays.
MOT2. I was willing to spend more time revising my writing.
MOT3. I felt more confident when revising my vocabulary.
MOT4. I would like to continue using ChatGPT in future writing courses.
MOT5. ChatGPT increased my interest in learning academic writing.

Appendix B. Standardized ChatGPT Feedback Protocol and Sample Interaction

B1. Standardized ChatGPT Feedback Protocol

To ensure consistency throughout the intervention, all participants in the experimental group followed the same ChatGPT feedback protocol. Students were instructed to write the first draft of each argumentative essay independently before interacting with ChatGPT. They were not permitted to ask ChatGPT to generate essays, create arguments, or rewrite entire paragraphs. Instead, ChatGPT was used exclusively as a lexical feedback tool during the revision stage.

All students used the same version of ChatGPT and submitted an identical prompt for each writing task. They were encouraged to evaluate every suggestion critically and decide independently whether to accept, modify, or reject the proposed revisions. The instructor monitored classroom implementation to ensure that the protocol was followed consistently across all participants.

B2. Standardized ChatGPT Prompt

Students entered the following prompt after completing the first draft of each essay:

Please review my argumentative essay.

1. Identify inappropriate, repetitive, or informal vocabulary.
2. Suggest more appropriate academic vocabulary where necessary.
3. Recommend more natural English collocations.
4. Suggest paraphrases while preserving my original meaning.
5. Briefly explain the reason for each suggestion.
6. Do not rewrite my entire essay or generate new arguments.
7. Focus only on improving lexical resource.

B3. Sample ChatGPT Feedback

Student's Original Sentence

Social media has many bad effects on university students.

ChatGPT Feedback

Aspect	ChatGPT Suggestion	Explanation
Vocabulary choice	Replace bad effects with negative consequences	"Negative consequences" is a more formal academic expression commonly used in argumentative writing.
Word precision	Replace many with numerous	"Numerous" is more precise and appropriate for academic writing.
Collocation	Use have negative consequences for instead of have bad effects on	This collocation is more natural in formal English.

Student's Revised Sentence

Social media may have numerous negative consequences for university students.

B4. Sample Paraphrasing Feedback

Student's Original Sentence

Students use mobile phones a lot during class.

ChatGPT Suggestion

Consider replacing use mobile phones a lot with spend excessive amounts of time using smartphones during class

Explanation

The revised expression is more precise, formal, and better suited to academic argumentative writing.

Student's Revision

Many students spend excessive amounts of time using smartphones during class.

B5. Sample Academic Vocabulary Enhancement

Student's Original Sentence

Many students think online learning is good.

ChatGPT Suggestion

Replace:

- *good → beneficial*
- *think → believe*
- *online learning → online education*

Explanation

The suggested vocabulary increases lexical sophistication while preserving the original meaning.

Student's Revision

Many students believe that online education is beneficial to university learning.

B6. Students' Responsibilities During Revision

Students were instructed to:

- complete the first draft independently before requesting feedback;
- review every ChatGPT suggestion critically;
- accept, modify, or reject suggestions according to their own judgment;
- maintain the original arguments and organization of the essay;
- avoid requesting complete essays, paragraph rewriting, or AI-generated content.

ChatGPT functioned solely as a lexical feedback provider rather than as a content generator.

B7. Instructor Monitoring

Throughout the intervention, the instructor monitored students' interactions with ChatGPT to ensure adherence to the research protocol. Students completed all writing tasks in class, and ChatGPT was used only during the designated revision sessions. The instructor also reminded students that the purpose of ChatGPT was to facilitate vocabulary refinement and lexical development rather than to replace independent writing or critical thinking.

Appendix C. Pre-test and Post-test Writing Tasks

Writing Task Instructions

Both the pre-test and post-test writing tasks were evaluated using a study-adapted analytic rubric informed by the IELTS Writing Task 2 Lexical Resource band descriptors. Four dimensions were assessed: Vocabulary Range, Word Choice Accuracy, Academic Vocabulary and Collocations, and Lexical Flexibility. Each dimension was rated on a nine-band scale (1–9). The final lexical resource composite score was calculated as the arithmetic mean of the four dimension ratings. The same scoring criteria were applied to both groups at both assessment points. The detailed rubric and scoring procedure are presented in Appendix D1.

Pre-test Writing Task

Topic

Should university students be allowed to use artificial intelligence tools such as ChatGPT when completing academic writing assignments?

Write an argumentative essay of 250-300 words expressing your opinion. Support your ideas with relevant reasons and examples.

Post-test Writing Task

Topic

Do the advantages of social media outweigh its disadvantages for university students?

Write an argumentative essay of 250-300 words expressing your opinion. Support your ideas with relevant reasons and examples.

Scoring Criteria

Both writing tasks were evaluated using the Lexical Resource criterion adapted from the IELTS Writing Band Descriptors. Four aspects of lexical performance were assessed:

- Range of vocabulary
- Accuracy of word choice
- Appropriate use of collocations and academic expressions
- Flexibility in paraphrasing and lexical variation

Two experienced EFL writing instructors independently scored all essays. Prior to marking, the raters participated in a calibration session to ensure consistent interpretation of the scoring criteria. Any discrepancies were discussed until agreement was reached, and the consensus scores were used for statistical analysis.

Appendix D1. Lexical Resource Assessment Rubric

The lexical resource of each essay was assessed using an analytic rubric adapted from the IELTS Writing Task 2 Lexical Resource Band Descriptors (British Council, IDP, & Cambridge Assessment English, 2023). The rubric was adapted to align with the objectives of the present study, which focused specifically on vocabulary development in EFL argumentative writing.

Four dimensions of lexical performance were evaluated:

1. Vocabulary range
2. Word choice accuracy
3. Academic vocabulary and collocations
4. Lexical flexibility (paraphrasing and lexical variation)

Each dimension was rated on a five-point scale ranging from 1 (very limited performance) to 9 (excellent performance). The overall lexical resource score for each essay was calculated by averaging the four dimension scores.

Table D1a.

Adapted IELTS Lexical Resource Analytic Scoring Rubric (Band 1–9)

Band	Vocabulary Range	Word Choice Accuracy	Academic Vocabulary and Collocations	Lexical Flexibility
9	Exceptionally wide, precise, and natural vocabulary	Consistently precise and appropriate lexical selection	Sophisticated, natural collocations and academic expressions	Highly flexible paraphrasing with negligible repetition
8	Wide vocabulary used fluently	Very accurate word choice; rare minor inaccuracies	Strong, natural collocations and academic register	Flexible variation and effective paraphrasing
7	Sufficient range for precise expression	Generally accurate with occasional inappropriate choices	Appropriate academic vocabulary and collocations	Effective paraphrasing with some repetition
6	Adequate range for the task	Meaning generally clear despite some imprecision	Some appropriate academic vocabulary and collocations	Attempts variation and paraphrasing with uneven success
5	Limited but usable range	Noticeable inappropriate word choices	Limited academic register and collocational control	Repetition and limited successful paraphrasing
4	Narrow range; frequent repetition	Frequent inappropriate choices impede clarity	Weak collocational and academic expression	Very limited variation
3	Very limited vocabulary	Frequent errors obscure meaning	Very few appropriate collocations	Minimal paraphrasing
2	Extremely restricted lexical repertoire	Persistent lexical errors	Almost no usable collocations	Little or no variation

1	Isolated words or formulaic expressions	Word selection largely ineffective	No sustained academic vocabulary use	No meaningful paraphrasing

Table D1b

Examples of Lexical Resource Performance Across Assessment Dimensions

Lexical Resource Dimension	Example of Low Performance	Example of High Performance
Vocabulary range	*good, bad, nice*	*beneficial, detrimental, advantageous*
Word choice	*bad effects*	*negative consequences*
Collocations	*do a decision*	*make a decision*
Paraphrasing	repeated *students*	*undergraduates, learners, university students*

Scoring Procedure

Each essay was independently evaluated by two experienced EFL writing instructors, both of whom had more than five years of experience teaching and assessing academic writing at the university level. Before the formal assessment, both raters participated in a calibration session using benchmark essays representing different levels of lexical performance. During this session, the raters discussed the scoring descriptors, compared their ratings, and established a shared interpretation of each assessment dimension.

The assessment employed an adapted analytic rubric informed by the IELTS Writing Task 2 Lexical Resource band descriptors. Four dimensions were evaluated: Vocabulary Range, Word Choice Accuracy, Academic Vocabulary and Collocations, and Lexical Flexibility. Each dimension was scored on a nine-band scale (1–9), with Band 0 reserved for responses containing no assessable language. These four dimensions represented the study's operationalization of lexical resource performance rather than four independently scored criteria in the official IELTS assessment system.

Each rater independently assigned scores for the four lexical dimensions. When differences between raters exceeded one band on any dimension, the ratings were discussed until consensus was reached. The agreed dimension scores were then used to calculate the final lexical resource score for each essay.

The final lexical resource score was calculated as the arithmetic mean of the four dimension scores:

Lexical Resource Score = (Vocabulary Range + Word Choice Accuracy + Academic Vocabulary and Collocations + Lexical Flexibility) / 4

For essays receiving assessable scores, the resulting composite ranged from 1.00 to 9.00. This composite was retained to two decimal places for statistical analysis. The scoring approach was adapted for the purposes of the present study and should not be interpreted as an official IELTS Writing Task 2 band score.

Sample Scoring Illustration

The following example illustrates how lexical resource was evaluated.

Student's Original Sentence

Social media has many bad effects on university students.

Evaluation

Dimension	Score	Justification
Vocabulary Range	**5**	Adequate vocabulary is demonstrated, but lexical variation is limited.
Word Choice Accuracy	5	The phrase *bad effects* is understandable but lacks lexical precision and academic appropriateness.
Academic Vocabulary and Collocations	5	Limited use of academic vocabulary and an unnatural lexical choice for formal argumentative writing.
Lexical Flexibility	5	Some variation is evident, although opportunities for paraphrasing remain limited.

Overall Lexical Resource Score

$$\frac{5+5+5+5}{4} = 5.0$$

Suggested Revision

Social media may have numerous negative consequences for university students.

Re-evaluation

Dimension	Score	Justification
Vocabulary Range	6	Greater lexical variety is demonstrated through more precise vocabulary choices.
Word Choice Accuracy	7	Academic vocabulary is accurate and contextually appropriate.
Academic Vocabulary and Collocations	7	The collocation *negative consequences* is natural and appropriate for argumentative writing.
Lexical Flexibility	6	The revised sentence demonstrates improved lexical sophistication while preserving the original meaning.

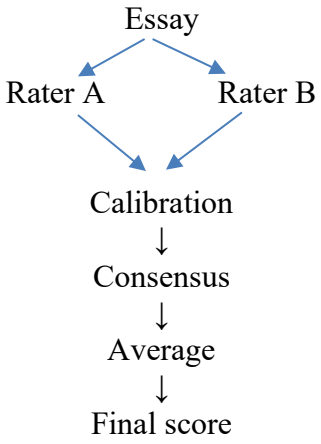
Revised Lexical Resource Score

$$\frac{6+7+7+6}{4} = 6.50$$

The improvement illustrates how lexical feedback encouraged students to replace informal expressions with more precise academic vocabulary, employ more natural collocations, and produce lexically richer argumentative writing while maintaining the intended meaning.

Figure D1.

Summary of the Lexical Resource Scoring Process



Appendix D2. Sample Scored Essays and Illustrative Lexical Resource Scoring

Sample Essay 1 (Approximately Band 5 for Lexical Resource)

Topic: *Should university students use social media in learning?*
Social media is **good** for students because they can learn many things from it. However, it also has many **bad effects**. Students spend too much time online and they become lazy. Many students use Facebook every day, and this is **bad** for study. Teachers should tell students to use social media carefully because it is **good** only when students use it correctly.

Lexical Resource Score: Band 5.0
Rater comments
Vocabulary range
- Frequent repetition of common words (*good, bad, students*).
- Limited use of academic vocabulary.

Word choice
- Meaning generally clear but vocabulary lacks precision.

Collocations
- Mostly acceptable but simple and repetitive.

Paraphrasing
- Little lexical variation.
- Few attempts to restate ideas using alternative expressions.

Sample Essay 2 (Approximately Band 7 for Lexical Resource)

Social media can be a **valuable educational resource** when it is used appropriately. It enables **undergraduates** to access learning materials efficiently and collaborate with their peers. Nevertheless, excessive use may result in **negative consequences**, including reduced concentration and academic procrastination. Universities should therefore encourage students to adopt responsible online learning practices while maintaining a healthy balance between digital engagement and independent study.

Lexical Resource Score: Band 7.0
Rater comments
Vocabulary range
- Wide range of academic vocabulary.

Word choice
- Vocabulary is precise and appropriate throughout.

Collocations
- Effective use of natural academic collocations (e.g., *valuable educational resource*, *negative consequences*, *responsible online learning practices*).

Paraphrasing
- Demonstrates lexical flexibility by avoiding repetition through expressions such as *undergraduates*, *students*, and *their peers*.