Opportunities and Challenges of Using Gemini AI for English Language Learning

[1]Jusak Patty, [2]Felicia Miranda Lekatompessy
[1,2]English Education Study Program, Pattimura University
Corresponding Email: jusak.patty@gmail.com

ABSTRACT

Generative AI (GenAI) tools, such as Gemini, offer new possibilities for English language learning; however, empirical evidence remains limited, particularly in resource-constrained contexts. This quantitative survey examined how 116 English Education undergraduates at Pattimura University in Eastern Indonesia perceived the usefulness of Gemini and the challenges they encountered in listening, speaking, reading, and writing skills. Students rated Gemini highly for all four skills, with reading receiving the strongest endorsement, followed by writing, speaking, and listening. Text-based features proved more reliable and accessible than audio-based functions. The most valued features included explaining complex vocabulary, improving sentence structure, generating presentation scripts, and creating transcripts from audio materials. However, significant challenges emerged. Academic integrity concerns dominated, with students uncertain about proper citation practices and worried that using AI suggestions might constitute plagiarism or reduce originality. Technical barriers, particularly internet connectivity and file format compatibility, hindered consistent access for many students. Information overload from lengthy responses also created difficulty selecting helpful feedback. The results underscore the importance of establishing clear institutional policies on AI collaboration, providing targeted training in the critical evaluation of AI output, and investing in infrastructure to ensure equitable access to AI. While students recognized Gemini's potential benefits, ethical uncertainties and technical constraints limited its optimal use, highlighting the importance of coordinated support across pedagogical, policy, and infrastructural dimensions.

Keywords: *Gemini AI, language learning, technology acceptance, academic integrity, English education*

INTRODUCTION

English language proficiency is a central requirement for Indonesian university students pursuing academic and professional goals. Listening, speaking, reading, and writing form the foundation of communicative competence and are assessed through both standardized tests and academic coursework (Guzmán-Simón et al., 2020; Nešić & Hamidovic, 2022). Yet, EFL learners at the tertiary level continue to face persistent challenges, particularly in productive skills, where opportunities for authentic practice outside the classroom remain scarce (Aisyah et al., 2024; Daflizar, 2020; Menggo et al., 2025). Many undergraduates enter English Education programs below the B2 CEFR benchmark needed for active academic participation (Baharum et al., 2021; Sae-Ong & Mohamed Ismail, 2021). These achievement gaps persist despite conventional pedagogical approaches, leading educators to consider digital tools that can provide individualized practice and real-time feedback.

GenAI has become one of the most significant technological innovations in education in recent years. Its ability to generate human-like responses, analyze complex input, and deliver instant feedback has positioned it as a promising support tool for language learning (Patty, 2024; Yan et al., 2024). Empirical studies consistently report positive outcomes, including improved writing performance (H. Li et al., 2025; Tiandem-Adamou, 2025), lower language anxiety (Rasheed, 2025; C. Wang et al., 2024), and stronger learner motivation (Leong et al., 2024; Renfeng et al., 2025). These findings suggest that GenAI can complement traditional pedagogies by offering more personalized learning opportunities and scalable support for diverse student needs. At the same time, the pace of technological development has already

moved beyond text-based interaction, producing multimodal models with expanded capabilities that extend to audio, visual, and interactive functions.

One of the most prominent examples of this new generation of tools is Google's Gemini AI, launched in December 2023. Unlike earlier models that relied primarily on text, Gemini integrates multimodal functions capable of processing audio, images, and video alongside written language (Imran & Almusharraf, 2024). This design opens possibilities for developing all four language skills: audio-based listening comprehension, voice interaction for speaking practice, text analysis for reading, and iterative feedback for writing. For Indonesian universities, Gemini offers particular advantages, including free access through Google Workspace for Education, seamless integration with platforms already familiar to students such as Drive and Classroom, and mobile accessibility that supports learners with limited device availability (Ananda & Salmiah, 2024; Imran & Almusharraf, 2024). Despite these affordances and nearly a year of public access, empirical studies investigating Gemini's effectiveness in educational contexts remain scarce. A recent systematic review of 144 studies on GenAI in language learning mentioned Gemini only in passing, underscoring the absence of direct evaluations of its outcomes (B. Li et al., 2025).

The few available studies provide preliminary but fragmented insights. Alnasib & Alharbi (2024) reported higher student motivation after the use of Gemini in Saudi classrooms. However, their design was limited by the absence of a control group and the reliance on self-reported perceptions rather than objective performance measures. Utami et al. (2024) conducted interviews with Indonesian undergraduates in North Sumatra, who reported greater confidence in their English learning after using Gemini; however, they also noted recurring technical issues, such as inaccurate translations, speech recognition errors, and unstable internet connections. Ananda & Salmiah (2024) reported that Indonesian students recognized the potential benefits of Gemini for the writing process but tended to favor the AI applications they were already accustomed to, reflecting the influence of established habits and the perceived costs of switching to a new tool. Nguyen et al. (2025) investigated Gemini's formative feedback on student essays in Vietnam and found that its comments aligned more closely with IELTS rubrics when prompts were explicitly structured. However, attention to surface-level features such as spelling and punctuation remained inconsistent. In a high school setting, Marpaung (2024) observed improvements in speaking scores over three classroom cycles; however, the single-group design limited the strength of causal interpretation.

Together, these studies reveal significant limitations. None have yet investigated all four language skills within a single framework; instead, evidence remains fragmented across isolated competencies or focuses more on motivational outcomes than performance measures. Small samples constrain generalizability, while accuracy concerns, unstable connectivity, and challenges in formulating prompts appear consistently. Academic integrity issues are often mentioned but rarely subjected to systematic measurement. Importantly, no research has been conducted in Eastern Indonesia, particularly Maluku province, where infrastructure conditions and student access patterns differ substantially from Western Indonesian contexts such as North Sumatra (Ananda & Salmiah, 2024; Marpaung, 2024; Utami et al., 2024), Saudi Arabia (Alnasib & Alharbi, 2024), and Vietnam (Luat et al., 2025). As a result, educators lack evidence-based guidance on how Gemini can be effectively integrated into local university settings across Indonesia's geographically diverse archipelago.

The present study addresses these gaps by investigating English Education undergraduates' experiences with Gemini AI at Pattimura University in Maluku, Eastern Indonesia. Grounded in technology acceptance principles, the study moves beyond descriptive accounts toward systematic analysis of students' perceptions of adoption and the challenges

they encounter. Using a quantitative survey, it examines the perceived usefulness and ease of use for developing listening, speaking, reading, and writing skills, while also assessing concerns about accuracy, academic integrity, and technical barriers. Two research questions guide the inquiry: (1) To what extent do students perceive Gemini AI as helpful and easy to use for developing the four language skills? (2) What challenges do students encounter when using Gemini for each skill?

This study situates Gemini within an Eastern Indonesian university context that has not been previously examined. By analyzing all four language skills together, it identifies where students see clear benefits and where they encounter barriers. The findings provide practical guidance for teachers on how Gemini can be utilized productively in the classroom and highlight the additional support learners may require. At the institutional level, the study also highlights the importance of infrastructure, training, and policy considerations in promoting sustainable adoption. In this way, the research provides evidence that extends beyond small case reports and informs AI integration in comparable resource-limited university settings.

LITERATURE REVIEW
Technology Acceptance Model in Educational AI Integration

Students decide whether to adopt new technologies through a series of cognitive judgments about performance benefits and effort requirements. Davis (1989) developed the Technology Acceptance Model (TAM) to capture this decision-making process, proposing that people ask themselves two fundamental questions before committing to regular technology use: Will this tool help me accomplish my goals more effectively? And will using it require more effort than I'm willing to invest? Educational researchers have found these questions reliably predict whether students will integrate digital tools into their learning routines (Sayaf et al., 2022; Songkram et al., 2023). What matters most is not the technology's objective capabilities but how individual users perceive its value and usability. The same learning platform might seem invaluable to one student yet frustratingly complex to another, depending on their prior experiences, task demands, and available support.

Applying TAM to AI in language classrooms has revealed both familiar patterns and unexpected complications. Several studies confirmed with Chinese university students that performance expectations drive adoption decisions, though ease of use must reach acceptable levels first (Liang et al., 2024; Y. Song et al., 2025). However, GenAI introduces challenges that are absent from conventional educational software. Students now worry not just about navigating interfaces but about crafting effective prompts and evaluating output reliability (Knoth et al., 2024; L. Wang et al., 2024). Karran et al. (2024) argue that trust becomes central when learners interact with systems that generate plausible yet potentially incorrect responses. Can I believe what this AI tells me? This epistemic concern extends beyond traditional usability considerations, suggesting researchers must account for how black-box algorithms shape user confidence alongside practical interaction costs.

GenAI Applications across Language Skills

Research documenting the impact of AI on writing and reading has produced encouraging results when interventions include structured feedback mechanisms. Writing studies consistently demonstrate skill improvements that persist beyond the immediate intervention period. Mahapatra (2024) tracked Indian undergraduates through multiple assessment points, finding that writing gains remained evident two months after AI support ended. This pattern suggests genuine skill development rather than temporary scaffolding dependence. C. Song & Song (2023) worked with Chinese students preparing for IELTS,

observing powerful improvements in how students organized their essays and developed coherent arguments. Reading research, although less developed, suggests that AI scaffolding helps students both understand texts more effectively and develop more sophisticated reading strategies. Shafiee Rad (2025) found that Iranian students not only comprehended passages more successfully but also monitored their understanding more actively, implying that AI support can cultivate transferable metacognitive skills.

Speaking and listening abilities have received surprisingly little research attention given their centrality to communication. Lo et al.'s (2024) review of 70 empirical studies found that no investigations examined listening comprehension with GenAI, while oral production appeared in only 7% of published research. The limited research available on speaking paints an incomplete picture. Mingyan et al. (2025) observed significant improvements among Chinese undergraduates who practiced with AI-powered applications; however, these gains were concentrated in pronunciation and fluency, while vocabulary and grammar showed minimal change. Interface design appears to matter considerably for speaking practice. C. Wang et al. (2024) discovered that adding visual avatars to AI conversational partners reduced anxiety more effectively than text-only interactions, highlighting how seemingly minor design choices influence learning outcomes. For listening, researchers have explored only adjacent technologies, such as speech recognition systems. Xiao (2025) tested such tools with positive results, but GenAI's unique capabilities for creating comprehension questions, producing scaffolded audio, or adapting input complexity await systematic investigation. This leaves educators without empirical guidance precisely where authentic practice opportunities prove scarcest outside formal instruction.

Implementation Challenges and Contextual Factors

GenAI use in language classrooms is not without problems, and one of the most common is the reliability of its output. Students expressed concerns about the accuracy of GenAI's production, particularly in academic contexts (Alnasib & Alharbi, 2024). While the study did not directly address cultural appropriateness, such issues are frequently raised in broader discussions of GenAI in language education. Outcomes also differ depending on proficiency. Etkin et al. (2025) found that weaker readers benefited from AI scaffolding, while stronger readers sometimes performed worse, showing that uniform use may disadvantage advanced learners. Concerns about academic integrity add another layer of complexity. Many students remain unsure how to acknowledge AI support or whether relying on it for revisions crosses ethical boundaries (Karkoulian et al., 2024). Over-reliance can also reduce independent thinking. Liu et al. (2025) reported that some learners became passive when they let AI produce answers instead of working through tasks themselves. These findings suggest that teachers should provide more explicit guidance, both on how to verify AI-generated content and on what constitutes acceptable collaboration (Patty et al., 2025). Training in prompt use and transparent classroom policies can help reduce these risks.

Infrastructure constraints also shape how students experience AI tools, especially in settings with limited resources. Utami et al. (2024) reported that Indonesian learners encountered unstable internet connections and other technical issues. Students adapted by scheduling their study more effectively, utilizing supplementary materials, and engaging in virtual learning groups. Bulathwela et al. (2024) reported that AI platforms with heavy bandwidth demands excluded students without suitable equipment, turning promising innovations into new inequalities. Such gaps are evident in archipelagic countries like Indonesia, where metropolitan areas enjoy stronger infrastructure than outlying regions. As Y. Li et al. (2025) argue, successful adoption depends not only on the technology itself but also

on the match between available infrastructure and classroom goals. Where internet access, devices, and institutional support are lacking, students are likely to face uneven opportunities to use AI effectively.

METHOD
Research Design
This study employed a quantitative survey design to examine English Education students' perceptions of Gemini AI across four language skills. The survey approach enabled systematic measurement of both opportunities and challenges within a single framework while maintaining ecological validity, as participants engaged with the technology under authentic coursework conditions rather than controlled experimental settings. This design suited the research objectives by documenting naturally occurring usage patterns after students had accumulated substantive experience through at least one whole semester of autonomous Gemini integration in their language learning activities.

Research Site and Participants
The research was conducted at the English Education Study Program, Pattimura University, in Ambon, Maluku Province. This Eastern Indonesian setting provided valuable context, given the absence of prior Gemini research in the region and the distinctive infrastructure conditions characterizing island universities, where connectivity and device access patterns differ substantially from those of Java-based institutions. The study population consisted of undergraduate English Education students enrolled during the 2024/2025 academic year who had integrated Gemini into their language learning coursework for at least one complete semester.

Convenience sampling was employed to recruit participants based on accessibility and willingness to participate during the data collection period. The final sample consisted of 116 students from the first through fifth year of study, offering perspectives shaped by varying English proficiency levels and academic task demands across intermediate to advanced stages of their teacher education program. As shown in Table 1, the sample composition revealed a pronounced gender imbalance characteristic of language education programs in Indonesia, with female students comprising 81.9% (n = 95) of the participants. In comparison, male students represented 18.1% (n = 21). This distribution aligns with documented enrollment patterns in English teaching programs at Indonesian universities, where female students have traditionally predominated.

The academic year distribution demonstrated a concentration in recent cohorts, with the 2023 intake contributing the largest proportion at 39.7% (n=46), followed by the 2024 intake at 31.9% (n=37). The 2022 cohort accounted for 18.1% (n = 21), while earlier cohorts showed minimal representation, with 2021 students at 7.8% (n = 9) and 2020 students at 2.6% (n = 3). This pattern reflected both the natural progression of the program, where senior students graduate or reduce their participation in voluntary surveys, and the relatively recent introduction of Gemini AI in December 2023, which meant that earlier cohorts had less exposure to the technology during their core coursework periods.

Duration of Gemini use varied considerably across the sample, indicating different adoption timelines and integration depths. Nearly half of the participants (48.3%, n = 56) reported three semesters of experience, suggesting early adoption shortly after Gemini's public release. Students with one semester of use comprised 28.4% (n = 33), representing more recent adopters or those who began systematic integration later in their academic progression. Those with two semesters of experience accounted for 23.3% (n = 27). This distribution provided

diverse perspectives spanning initial experimentation through sustained long-term use, enabling the study to capture perceptions across different stages of technology familiarity and skill development. Students who used Gemini only sporadically or primarily for non-academic purposes were excluded to maintain focus on educational applications and ensure respondents possessed substantive experience with the technology in language learning contexts.

Table 1. Demographic Profile of Participants

Demographic Variable	Category	n	%
Gender	Male	21	18.1
	Female	95	81.9
Academic Cohort	2020	3	2.6
	2021	9	7.8
	2022	21	18.1
	2023	46	39.7
	2024	37	31.9
Duration of Gemini Use	1 Semester	33	28.4
	2 Semesters	27	23.3
	≥ 3 Semesters	56	48.3

Data Collection and Analysis

The research instrument consisted of two primary dimensions that measured opportunities and challenges across listening, speaking, reading, and writing skills. The Opportunities dimension assessed perceived usefulness and perceived ease of use through 16 items, with each skill receiving four items addressing performance enhancement beliefs and interaction effort requirements. The Challenges dimension captured accuracy concerns, academic integrity issues, and technical barriers through 12 items, with each skill domain receiving three items reflecting specific obstacles students encountered. All 28 items employed a five-point Likert scale, ranging from "strongly disagree" to "strongly agree", using the Indonesian language to ensure clarity of comprehension and reduce interpretation errors among respondents.

Before the main data collection, pilot testing was conducted with 30 English Education students who were not included in this study's participants. The pilot phase was used to assess item clarity, identify ambiguous wording, evaluate respondent comprehension of the instructions, and test the technical functionality of the online survey platform. Participants provided feedback on the understandability of items and the time required to complete the survey, which informed minor revisions to phrasing and format before finalizing the instrument for the main study.

Instrument validity was established through content validity procedures, with all 28 items demonstrating adequate validity coefficients and retained for the final instrument. Reliability analysis using pilot testing data yielded Cronbach's alpha values exceeding the acceptable threshold of 0.70 for both the Opportunities dimension and Challenges dimension, confirming internal consistency of the measurement scales. These psychometric properties ensured the instrument reliably measured the intended constructs across the four language skill domains.

Data collection spanned four weeks during the mid-semester period, utilizing Google Forms, which was chosen for its device compatibility and ease of use. These offline-capable features addressed connectivity constraints and interface familiarity among Indonesian students. The survey began with informed consent procedures that explained the research purposes, voluntary participation, withdrawal rights, and confidentiality protections before participants accessed the questionnaire items. Researchers remained available throughout the collection period to address technical difficulties and clarify item meanings, sending reminder

messages that encouraged participation while respecting the principles of voluntary engagement. Survey completion required approximately 20 to 25 minutes, with automatic response saving enabling students to complete the questionnaire across multiple sessions when necessary.

Analysis focused on descriptive examination of response patterns to address the research questions. Descriptive statistics summarized the central tendency and variability for each construct across the four language skills, with mean scores interpreted using standardized intervals to determine perception levels. Standard deviations and frequency distributions complemented mean score analysis by revealing response variability and distribution patterns within each construct. Skill-specific opportunity and challenge profiles emerged by organizing data within each language domain, creating portraits that identified where students found Gemini most useful and easiest to use, as well as which accuracy concerns, academic integrity issues, and technical barriers proved most salient for each skill area. The interpretation criteria employed are presented in Table 2.

Table 2. Mean Score Interpretation Criteria

Mean Score Range	Interpretation Level
1.00 – 1.80	Very Low
1.81 – 2.60	Low
2.61 – 3.40	Moderate
3.41 – 4.20	High
4.21 – 5.00	Very High

FINDINGS AND DISCUSSION
Research Findings

This section presents the findings that address the two research questions guiding this investigation. First, students' perceptions of Gemini AI's usefulness and ease of use are examined across the four language skills. Second, the challenges students encountered when using Gemini for developing each skill are analyzed. Descriptive statistics, including means, standard deviations, and frequency distributions, provide a systematic portrait of student experiences with this multimodal AI tool in an Eastern Indonesian university context.

Perceived Usefulness and Ease of Use across Language Skills

Students held positive views of Gemini AI across all four language skills, as presented in Table 3. The overall mean reached 3.70 (SD = 0.80), falling within the "High" range based on the interpretation criteria. After using Gemini for one to three semesters, most students found the tool both helpful and accessible for their language learning activities. The four skills showed different levels of acceptance. Reading received the highest rating, at 3.85 (SD = 0.75), followed by writing, at 3.77 (SD = 0.77). Speaking came next at 3.64 (SD = 0.82), while listening scored the lowest at 3.52 (SD = 0.83), although it was still in the high category. Students appeared most comfortable with text-based tasks, such as reading comprehension and essay writing, where they could analyze and compose at their own pace. Audio-based features for listening practice presented more difficulties, possibly due to technical constraints or a lack of familiarity with these functions.

Table 3. Summary of Perceived Usefulness & Ease of Use Means

Language Skill	Mean	SD	Category
Reading	3.85	0.75	High
Writing	3.77	0.77	High

Speaking	3.64	0.82	High
Listening	3.52	0.83	High
Overall	**3.70**	**0.80**	**High**

1. Listening Skills

Requesting transcripts or summaries emerged as the most valued listening feature (L4, M = 3.71, SD = 0.90). Eighty students (69.0%) found this helpful, with 63 selecting "agree" and 17 selecting "strongly agree." Only 12 students disagreed. The transcript function helped reduce mental load when processing spoken English, particularly for academic lectures where students needed to capture specific details. Twenty-four students remained neutral, possibly because they had not yet had the opportunity to try this feature extensively.

Getting keywords and outlines for note-taking (L2, M=3.50, SD=0.84) also received positive responses from 64 students (55.2%). Similarly, 57 students (49.1%) appreciated how Gemini helped them identify main ideas in audio materials (L1, M = 3.47, SD = 0.71). Both features had many neutral responses, though, with 42 students for L2 and 53 for L1. This suggests that some students either had limited experience with these tools or remained unsure about their effectiveness. Preparing listening exercises by uploading audio files or links (L3, M = 3.41, SD = 0.82) received the lowest rating among the listening items. Fifty-four students (46.6%) found it helpful, but 51 chose the neutral option. The high neutral count suggests potential technical issues. Students may have struggled with file format compatibility or internet connectivity when trying to upload audio materials. The complete distribution of responses appears in Table 4.

Table 4. Perceived Usefulness and Ease of Use for Listening Skills

No	Item	Response Distribution					Mean	SD	Category
		1	2	3	4	5			
L4	I can easily request transcripts or summaries of key points from audio with simple prompts	3	9	24	63	17	3.71	0.90	High
L2	Gemini improves my note-taking skills from audio by providing keywords and brief outlines	4	6	42	56	8	3.50	0.84	High
L1	Gemini helps me capture the main ideas from English audio, such as lectures or news	2	4	53	52	5	3.47	0.71	High
L3	I can easily prepare listening exercises in Gemini by uploading links or audio and requesting tasks	3	8	51	46	8	3.41	0.82	High
						Overall Mean	**3.52**	**0.83**	**High**

2. Speaking Skills

Creating practice scripts for presentations (S4) stood out as the most valued speaking feature, with a mean of 3.78 (SD = 0.89). Seventy-two students (62.1%) responded positively, including 26 who strongly agreed. This was the highest "strongly agree" count among all speaking items. Students clearly appreciated having structured frameworks for their presentations, which saved preparation time while maintaining quality. Only six students disagreed, and 38 remained neutral.

Starting conversation practice with simple instructions about roles and situations (S3, M = 3.66, SD = 0.78) was effective for 71 students (61.2%), with only five expressing disagreement. Getting feedback on expressions and intonation (S2, M=3.58, SD=0.80) helped 66 students (56.9%) make their speech sound more natural. Role-play exercises for fluency practice (S1, M = 3.54, SD = 0.80) received positive responses from 63 students (54.3%), while 46 selected a neutral response. Nearly 40% remaining neutral suggests that many students either had not experimented with role-play functions or found them less engaging than other

features. Role-play may not simulate honest conversations as effectively as practicing with actual people. Table 5 provides a detailed breakdown of the speaking responses.

Table 5. Perceived Usefulness and Ease of Use for Speaking Skills

No	Item	Response Distribution					Mean	SD	Category
		1	2	3	4	5			
S4	I can easily request brief practice scripts for presentation preparation	2	4	38	46	26	3.78	0.89	High
S3	I can easily start conversation practice with brief instructions about roles, situations, and duration	2	3	40	58	13	3.66	0.78	High
S2	Gemini is useful for providing feedback on expressions and intonation that make speech more natural	2	6	42	55	11	3.58	0.80	High
S1	Gemini helps improve my speaking fluency through role-play exercises on academic topics	3	4	46	53	10	3.54	0.80	High
					Overall Mean		**3.64**	**0.82**	**High**

3. Reading Skills

Explaining difficult vocabulary and complex sentences (R4) scored highest across all items in the survey, with a mean of 3.92 (SD = 0.78), as shown in Table 6. Eighty-eight students (75.9%) found this helpful, with only three expressing any disagreement. The concentration in positive responses (64 selecting "agree" and 24 selecting "strongly agree") shows strong consensus. Students clearly needed help with technical terminology and complex sentence structures in academic English, and Gemini effectively met this need.

Creating comprehension questions from uploaded texts (R3, M = 3.86, SD = 0.73) received similar endorsement from 87 students (75.0%). Generating structured summaries (R1, M = 3.85, SD = 0.73) was effective for 88 students (75.9%). Teaching reading strategies, such as skimming and making inferences (R2, M = 3.75, SD = 0.74), helped 77 students (66.4%). Disagreement remained minimal across all reading items, with a range of one to four students per item. Neutral responses were also relatively limited, between 23 and 35 students. Most students had direct experience with these reading features and recognized their value. The "agree" category dominated responses, indicating that text analysis support was consistently adequate across different student needs.

Table 6. Perceived Usefulness and Ease of Use for Reading Skills

No	Item	Response Distribution					Mean	SD	Category
		1	2	3	4	5			
R4	I can easily ask Gemini to explain difficult parts, such as technical terms or complex sentences	2	1	25	64	24	3.92	0.78	High
R3	I can easily ask Gemini to create comprehension questions from texts I upload	1	3	25	69	18	3.86	0.73	High
R1	Gemini helps me understand academic texts through structured summaries such as objectives, methods, and results	1	4	23	71	17	3.85	0.73	High
R2	Gemini helps teach reading strategies such as skimming, scanning, and making inferences in academic texts	1	3	35	62	15	3.75	0.74	High
					Overall Mean		**3.85**	**0.75**	**High**

4. Writing Skills

Sentence-level improvements (W4, M=3.89, SD=0.74) topped the writing features, with 81 students (69.8%) responding positively, including 24 who strongly agreed. No student strongly disagreed, and only two expressed mild disagreement. Students valued how Gemini refined their language without changing their original ideas. This addressed a common worry that AI might take over their personal writing voice.

Revision help aligned with instructor rubrics (W2, M = 3.78, SD = 0.69) was practical for 81 students (69.8%). No students strongly disagreed, and only four expressed disagreement. Getting outlines and essay structures (W3, M=3.76, SD=0.83) helped 75 students (64.7%), with 20 strongly agreeing. Paragraph coherence and organization (W1, M=3.66, SD=0.80) had the most neutral responses among writing items at 43 students. The approximately 37% remaining neutral suggests that working on the overall essay structure proved more challenging than fixing individual sentences. Students may have wanted more detailed guidance on argument development. Disagreement remained minimal across writing items, with zero to three students per item, indicating that Gemini's writing support effectively addressed real student needs. The detailed response patterns are presented in Table 7.

Table 7. Perceived Usefulness and Ease of Use for Writing Skills

No	Item	Response Distribution					Mean	SD	Category
		1	2	3	4	5			
W4	I can easily request suggestions for sentence improvements without changing the main ideas	0	2	33	57	24	3.89	0.74	High
W2	Gemini helps with revision, such as transitions, topic sentences, and example appropriateness according to instructor rubrics	0	4	31	67	14	3.78	0.69	High
W3	I can easily request outlines and essay structures according to given assignments	2	3	36	55	20	3.76	0.83	High
W1	Gemini helps improve the coherence and organization of paragraphs in my essays	2	3	43	53	15	3.66	0.80	High
					Overall Mean		3.77	0.77	**High**

Challenges Encountered across Language Skills

Students reported experiencing real challenges alongside the benefits, as presented in Table 8. The overall challenge mean reached 3.44 (SD = 0.87), still within the "High" range. These difficulties were not just minor inconveniences but persistent problems that affected how students used Gemini. The challenge score being lower than the 3.70 benefit score makes sense. Students saw value in Gemini, but they also faced obstacles that limited their ability to utilize it effectively.

Table 8. Summary of Overall Challenges Means

Language Challenges	Mean	SD	Category
Writing	3.56	0.83	High
Speaking	3.46	0.87	High
Listening	3.40	0.87	Moderate
Reading	3.32	0.88	Moderate
Overall	**3.44**	**0.87**	**High**

Different skills presented different challenge levels. Writing showed the highest challenge score, at 3.56 (SD = 0.83), followed by speaking, at 3.46 (SD = 0.87). Reading had a moderate challenge level of 3.32 (SD = 0.88), while listening fell between moderate and high,

at 3.40 (SD = 0.87). Reading showed an interesting pattern. It had the highest perceived usefulness, but still faced notable challenges regarding accuracy and citation. Listening showed lower benefits and moderate difficulties, suggesting that this skill area had problems with both practical use and achieving good learning results.

1. Listening Challenges

Technical problems dominated the listening challenges, as shown in Table 9. Internet connectivity and audio upload issues (LC3, M = 3.58, SD = 0.96) created the biggest obstacle. Sixty-five students (56.0%) agreed or strongly agreed that these problems hindered their practice. Nineteen students strongly agreed, the highest "strongly agree" count for any listening challenge. Technical barriers seriously disrupted learning for this group. Only 14 students reported that connectivity was not a problem, while 37 remained neutral. About two-thirds of students faced at least occasional technical difficulties. This reflects the infrastructure limitations in Eastern Indonesia, where bandwidth and network stability vary widely.

Table 9. Challenges in Listening Skills

No	Item	Response Distribution					Mean	SD	Category
		1	2	3	4	5			
LC3	Uploading audio or links and internet connectivity often hinder my listening practice	3	11	37	46	19	3.58	0.96	High
LC2	I am uncertain how to acknowledge AI assistance when writing reflections from audio materials	1	12	56	42	5	3.33	0.75	Moderate
LC1	Transcripts or audio summaries from Gemini are sometimes inaccurate and mislead my comprehension	2	17	47	43	7	3.31	0.86	Moderate
						Overall Mean	3.40	0.87	Moderate

Uncertainty about acknowledging AI help (LC2, M = 3.33, SD = 0.75) troubled 47 students (40.5%). The majority, 56 students (48.3%), selected neutral, showing widespread confusion about proper attribution. Only 13 felt confident they knew the rules. The lack of clear institutional guidance on citing AI assistance affected most students. Accuracy problems with transcripts and summaries (LC1, M = 3.31, SD = 0.86) were reported by 50 students (43.1%). Nearly half, 47 students (40.5%), stayed neutral. This could mean that accuracy issues occurred intermittently rather than consistently. Students may also have found it challenging to judge transcript reliability without reference materials to compare against.

2. Speaking Challenges

Authenticity concerns dominated the speaking challenges. Worries about losing originality (SC2, M = 3.76, SD = 0.94) emerged as the strongest concern across all challenge items. Seventy-two students (62.1%) expressed concern that relying on AI scripts might compromise their originality. Twenty-eight strongly agreed, the highest "strongly agree" count in the entire survey. This issue particularly matters for future English teachers who need to develop authentic personal expression and model natural language for their students. Only 11 students disagreed, while 33 remained neutral.

Unnatural or contextually inappropriate phrases (SC1, M = 3.39, SD = 0.80) bothered 55 students (47.4%). Nearly half, 46 students (39.7%), selected neutral. This suggests that quality problems sometimes occur, but not consistently. Students learned they needed to evaluate and modify AI suggestions rather than accept them automatically. Speech recognition and latency

problems (SC3, M = 3.23, SD = 0.78) disrupted practice for 41 students (35.3%). The majority, 60 students (51.7%), stayed neutral, while 15 disagreed. Technical issues affected a minority intensely, while most experienced only occasional disruptions. The lower technical challenge score compared to originality concerns shows that ethical and pedagogical questions mattered more to students than purely functional problems. These patterns are detailed in Table 10.

Table 10. Challenges in Speaking Skills

No	Item	Response Distribution					Mean	SD	Category
		1	2	3	4	5			
SC2	I worry that depending on Gemini scripts during presentations reduces my originality	1	10	33	44	28	3.76	0.94	High
SC1	Suggested spoken phrases from Gemini are sometimes unnatural or inappropriate for the speaking context	1	14	46	49	6	3.39	0.80	Moderate
SC3	Speech recognition and latency issues disrupt real-time conversation practice	3	12	60	37	4	3.23	0.78	Moderate
					Overall Mean		**3.46**	**0.87**	**High**

3. Reading Challenges

Technical format issues (RC3, M = 3.47, SD = 0.91) presented the most pressing reading challenge. Fifty-five students (47.4%) found that PDF compatibility and text length limits restricted their analysis. Forty-five students (38.8%) stayed neutral, suggesting these constraints affected students selectively based on their specific coursework requirements. Only 16 disagreed. Format restrictions often prevented students from comprehensively analyzing longer articles or book chapters.

Citation confusion (RC2, M = 3.27, SD = 0.91) affected 52 students (44.8%), who felt uncertain about citing AI help in reading summaries. The 41 neutral responses (35.3%), combined with 23 students who disagreed, show varied experiences. Different instructors may have given different guidance. Some students received clear citation rules, while others did not. Interpretive accuracy problems (RC1, M = 3.22, SD = 0.79) represented the lowest challenge score but still posed a significant issue. Thirty-nine students (33.6%) said Gemini's explanations sometimes contained wrong inferences or exaggerated claims. Most students, 59 (50.9%), selected the neutral option. Accuracy problems seemed to happen occasionally rather than systematically. Students learned to cross-check sources and evaluate AI interpretations critically instead of trusting them completely. Table 11 provides the complete data of reading challenges.

Table 11. Challenges in Reading Skills

No	Item	Response Distribution					Mean	SD	Category
		1	2	3	4	5			
RC3	File formats such as PDF or scans, and text length limits often restrict my reading analysis process	1	15	45	39	16	3.47	0.91	High
RC2	I am confused about when and how to cite AI assistance when creating reading summaries	4	19	41	46	6	3.27	0.91	Moderate
RC1	Gemini's text explanations sometimes have incorrect inferences or exaggerate reading claims	1	17	59	33	6	3.22	0.79	Moderate
					Overall Mean		**3.32**	**0.88**	**Moderate**

4. Writing Challenges

Plagiarism worries (WC2, M=3.68, SD=0.76) and information overload (WC3, M=3.66, SD=0.86) emerged as the two most substantial writing challenges, as detailed in Table 12. Sixty-eight students (58.6%) expressed concern that using Gemini's editorial suggestions might border on plagiarism or over-editing. Fifty-two selected "agree," and 16 "strongly agree," showing widespread anxiety. This went beyond citation confusion to fundamental questions about whether using AI suggestions counted as cheating. Only five disagreed. The lack of clear institutional policies left students uncertain about acceptable AI collaboration.

Managing verbose feedback (WC3) was a source of trouble for 71 students (61.2%). Gemini's thorough responses sometimes overwhelmed students who struggled to distinguish between beneficial suggestions and those that were less important. Fifty-four selected "agree," and 17 "strongly agree." Only nine disagreed, while 36 stayed neutral. Most students faced this challenge at least occasionally. Verifying references and examples (WC1, M $= 3.33$, SD $= 0.82$) was a concern with 48 students (41.4%). Most students, 53 (45.7%), selected 'neutral', suggesting that verification challenges occurred sometimes but not always. Only 15 disagreed. Most students recognized they needed to verify AI-provided information rather than accept it automatically, which reduced the efficiency gains from using AI assistance.

Table 12. Challenges in Writing Skills

No	Item	Response Distribution					Mean	SD	Category
		1	2	3	4	5			
WC2	I worry that editorial suggestions from Gemini border on plagiarism or over-editing	0	5	43	52	16	3.68	0.76	High
WC3	Lengthy or repetitive responses make it difficult for me to select vital suggestions	2	7	36	54	17	3.66	0.86	High
WC1	References or examples provided by Gemini are sometimes inaccurate or difficult to verify	2	13	53	41	7	3.33	0.82	Moderate
					Overall Mean		3.56	0.83	High

Discussion

English Education students at Pattimura University generally accept Gemini AI as a valuable and accessible tool for language learning. This acceptance aligns with the TAM's core premise that perceived usefulness and ease of use drive adoption decisions (Davis, 1989; Sayaf et al., 2022). Students who found Gemini helpful for specific tasks and easy to operate were more likely to integrate it into their learning routines, echoing patterns observed in other educational technology adoption studies (Songkram et al., 2023). The positive reception across all four skills indicates that, after sustained use, most students had progressed beyond initial experimentation to recognize concrete benefits. This progression mirrors the motivational gains reported by Alnasib & Alharbi (2024) in Saudi classrooms, where students demonstrated a greater willingness to participate after the classroom integration of Gemini. However, reading and writing received higher ratings than speaking and listening, indicating that text-based interactions were more suitable for students than audio-based features. This pattern differs from conventional language learning priorities, where oral skills often receive greater emphasis, particularly in communicative language teaching approaches that emphasize interpersonal communication (Menggo et al., 2025; Nešić & Hamidović.

The preference for reading and writing support reflects both the nature of academic coursework and the current maturity of AI capabilities. Students spend considerable time analyzing academic texts and composing essays, making Gemini's text processing functions

directly relevant to their immediate needs. The ability to explain complex vocabulary addresses a persistent challenge in academic English where discipline-specific terminology creates comprehension barriers (Kim et al., 2025; Y. Liu et al., 2024). Similarly, sentence-level improvements met a widely recognized need for linguistic refinement without sacrificing authorial voice. Recent studies show similar patterns, with AI-assisted writing improving both performance and motivation when students receive targeted feedback on specific aspects of composition (H. Li et al., 2025; Mahapatra, 2024; C. Song & Song, 2023). Luat et al. (2025) found that Gemini's feedback aligned more closely with IELTS rubrics when prompts explicitly included assessment criteria, though consistency varied across different writing dimensions. The sustained improvements reported by Mahapatra (2024) two months after intervention ended suggest genuine skill development rather than temporary scaffolding dependence. The lower ratings for listening features stem partly from technical constraints. Many students struggled with uploading audio files and connectivity issues, which limited their ability to utilize the transcript generation and summary functions fully. Utami et al. (2024) documented technical disruptions among North Sumatran learners, where unreliable internet access created uneven learning experiences, leading students to adopt adaptive strategies such as organizing their study schedules, joining virtual learning groups, and supplementing their learning with alternative materials. Unlike text-based tasks that work reliably across devices and connection speeds, audio processing demands stable bandwidth and compatible formats. The infrastructure limitations in Maluku created uneven access, leaving some students unable to use listening features consistently enough to judge their effectiveness.

The challenge profile reveals a different story. Writing and speaking presented higher challenges than reading and listening, inverting the usefulness pattern. Students found productive skills more problematic than receptive ones, though for different reasons. Writing challenges centered on academic integrity concerns, with most students worried about plagiarism or over-editing. This anxiety reflects the ambiguity surrounding the acceptable use of AI in educational contexts (Karkoulian et al., 2024). Without clear institutional policies, students remained uncertain whether incorporating Gemini's suggestions constituted improper assistance. Ananda & Salmiah (2024) identified a similar paradox among Indonesian undergraduates who acknowledged Gemini's value for writing yet resisted adoption due to established habits with other AI tools and uncertainty about proper usage boundaries. The struggle with information overload points to a practical problem. Gemini's thorough responses, while detailed, sometimes overwhelmed students who lacked strategies for filtering essential feedback from peripheral suggestions. J. Liu et al. (2025) observed similar patterns, noting that AI assistance can lead to reduced independent thinking when learners become passive recipients rather than active evaluators of the output. Luat et al. (2025) indicated that while Gemini provided consistent feedback on task achievement and coherence, it tended to neglect mechanical features, such as spelling and punctuation, creating asymmetries in coverage that required human oversight. The challenge becomes particularly acute when students face assignment pressures and may accept AI suggestions uncritically rather than engaging in the deliberate practice needed for skill development.

Speaking challenges took a different form. The strongest concern was losing originality, with intensity exceeding all other challenge items. This worry carries special weight for future English teachers who must develop authentic expression and model natural language for their students. The fear that AI-generated scripts might compromise personal voice reveals tension between efficiency and authenticity. Students appreciated having presentation frameworks, but worried about depending on them too heavily. This concern is connected to broader debates about AI's role in creative and expressive tasks, where individual voice matters as much as

technical accuracy (Doshi & Hauser, 2024; C. Wang, 2025; C. Wang et al., 2024). Marpaung (2024) reported progressive improvements in speaking scores across three classroom action research cycles with high school students; however, the single-group design raised questions about whether the gains reflected genuine skill development or novelty effects. C. Wang et al (2024) found that adding visual avatars to AI conversational partners reduced anxiety more effectively than text-only interactions, showing that interface design choices influence not just usability but also learners' comfort with AI-mediated language production. Utami et al. (2024) documented how speech recognition errors and connectivity disruptions forced students to adapt by using more formal prompts and seeking peer support in virtual groups. The relatively lower concern about technical issues compared to originality shows that students prioritized pedagogical questions over functional problems. They could tolerate occasional speech recognition errors or latency, but they could not accept losing their distinctive way of expressing ideas. This hierarchy of concerns shows that successful AI integration for speaking practice must address identity and authenticity issues before optimizing technical performance.

Academic integrity emerged as a cross-cutting concern affecting multiple skills. Beyond the plagiarism concerns of writing, students felt uncertain about acknowledging AI help in listening reflections and were confused about citing AI assistance in reading summaries. These parallel concerns reveal systemic gaps in institutional guidance rather than isolated uncertainties (Ateeq et al., 2024; Balalle & Pannilage, 2025). The variation in citation confusion shows that different instructors provided inconsistent direction, leaving students to navigate ethical boundaries independently. Luat et al. (2025) observed that explicit prompting structures improve the consistency and quality of AI output, particularly when rubric descriptors are included in the input. By analogy, just as AI requires clear prompts to perform effectively, students also need explicit institutional policies to guide their appropriate use of AI. Alnasib & Alharbi (2024) found that nearly half of Saudi students expressed uncertainty about the factual accuracy of their work, despite reporting motivational gains, revealing how enthusiasm coexists with caution regarding content quality and academic integrity. The absence of such guidance creates anxiety that undermines the learning benefits AI might otherwise provide (Knoth et al., 2024). Students spent cognitive resources worrying about rule violations rather than focusing on skill development. The challenge of AI literacy extends beyond technical prompt engineering to encompass ethical awareness and metacognitive judgment about when and how to leverage AI assistance appropriately (Kong et al., 2024). Ananda & Salmiah (2024) observed that even when students recognized Gemini's capabilities, behavioral inertia and uncertainty about acceptable use delayed actual adoption, showing that institutional clarity and training could catalyze behavior change. Universities must develop comprehensive frameworks that address not just citation mechanics but the broader questions of intellectual ownership and learning authenticity in AI-augmented environments.

The Eastern Indonesian context significantly influenced these findings. Technical barriers affected many students across listening and reading tasks, reflecting the realities of Maluku's infrastructure, where internet connectivity varies widely between urban centers and outlying areas. These rates appear higher than those reported in studies from other Indonesian universities (Ananda & Salmiah, 2024) or from international contexts with stronger digital infrastructure (Alnasib & Alharbi, 2024). Utami et al. (2024) documented remarkably similar challenges among North Sumatran learners, where inconsistent translation accuracy, speech recognition failures, and unequal internet access created disruptions that required students to devise organic adaptive strategies rather than relying on formal institutional support. The geographic isolation that characterizes archipelagic regions presents particular challenges for technology integration, which metropolitan-focused research often overlooks (Bulathwela et

al., 2024). Students adapted by scheduling their Gemini use around connectivity windows or focusing on offline-capable features, mirroring the coping strategies documented by Utami et al. (2024) among North Sumatran learners who used supplementary materials and virtual study groups to compensate for technical limitations. These workarounds, while demonstrating student resilience, still limit optimal use. Marpaung (2024) worked within a high school classroom setting, where controlled implementation may have mitigated some infrastructure challenges. However, the broader pattern across Indonesian studies reveals that technical barriers remain consequential, regardless of educational level. Successful AI integration depends not just on tool design but on the match between technological requirements and local infrastructure capacity (Y. Li et al., 2025). Bulathwela et al. (2024) warn that AI platforms with heavy bandwidth demands risk excluding students without suitable equipment, turning promising innovations into new inequalities. Policymakers and educators must consider these contextual factors when promoting AI adoption, ensuring that enthusiasm for innovation does not outpace the practical conditions necessary for effective implementation. The study also reveals that despite infrastructure challenges, students still found value in Gemini, suggesting resilience and adaptability in leveraging available resources for language learning improvement even when conditions fall short of ideal implementation scenarios.

CONCLUSION

English Education students at Pattimura University found Gemini AI to be generally helpful and accessible across four language skills, with reading and writing receiving higher acceptance ratings than speaking and listening. Text-based features worked more reliably and met academic needs better than audio-based functions. However, challenges emerged around academic integrity and originality, particularly for writing and speaking. Students worried whether using AI suggestions constituted plagiarism and feared losing their authentic voice. Technical barriers also limited access, particularly for listening tasks, as connectivity issues hindered consistent use. The results have practical implications for implementation. Teachers need clear guidance on acceptable AI collaboration to share with students. Without institutional policies on citation and proper use, anxiety about violations undermines learning benefits. Training can help students evaluate AI output critically rather than accept it passively. Infrastructure investment matters for equity, as uneven internet access in Eastern Indonesia created different experiences across locations. Theoretically, the study confirms that usefulness and ease of use drive adoption, but challenges can coexist with positive perceptions. Acceptance models must account for both ethical and contextual barriers, as well as technical factors.

Several limitations affect the interpretation of these results. The convenience sample from one university restricts generalizability to other regions. Self-reported perceptions may not reflect actual learning gains, as performance was not measured. The cross-sectional design captured views at a single point in time but could not track changes over time. Future research can employ longitudinal designs that track both perceptions and performance, experimental studies comparing AI-assisted and traditional learning, investigations across diverse Indonesian regions to test generalizability, and qualitative studies examining how students navigate academic integrity dilemmas. Research on instructor perspectives and institutional readiness will also help, as successful integration requires coordinated support. Building on this foundation, future work can guide evidence-based policies that maximize benefits while managing the real challenges students face.

REFERENCES

Aisyah, S., Rasyid, M. N. A., Nasrum, Djamereng, J., & Asnur, S. M. (2024). Exploring EFL Students' Perspective in Improving their English Productive Skills: Existing Competencies and Encountered Challenges. *Elite : English and Literature Journal*, *11*(1), 33–51. https://doi.org/10.24252/elite.v11i1.47166

Alnasib, B. N. M., & Alharbi, N. S. (2024). Challenges and Motivation: Assessing Gemini's Impact on Undergraduate EFL Students in Classroom Settings. *World Journal of English Language*, *14*(5), 501. https://doi.org/10.5430/wjel.v14n5p501

Ananda, D. R., & Salmiah, M. (2024). Students' Perception on AI Technology: Gemini as a Writing Assistant Tool. *Linguistics and ELT Journal*, *12*(1), 46. https://doi.org/10.31764/leltj.v12i1.24393

Ateeq, A., Alzoraiki, M., Milhem, M., & Ateeq, R. A. (2024). Artificial intelligence in education: Implications for academic integrity and the shift toward holistic assessment. *Frontiers in Education*, *9*, 1470979. https://doi.org/10.3389/feduc.2024.1470979

Baharum, N. N., Ismail, L., Nordin, N., & Razali, A. B. (2021). Aligning a University English Language Proficiency Measurement Tool with the CEFR: A Case in Malaysia. *Pertanika Journal of Social Sciences and Humanities*, *29*(S3). https://doi.org/10.47836/pjssh.29.s3.09

Balalle, H., & Pannilage, S. (2025). Reassessing academic integrity in the age of AI: A systematic literature review on AI and academic integrity. *Social Sciences & Humanities Open*, *11*, 101299. https://doi.org/10.1016/j.ssaho.2025.101299

Bulathwela, S., Pérez-Ortiz, M., Holloway, C., Cukurova, M., & Shawe-Taylor, J. (2024). Artificial Intelligence Alone Will Not Democratise Education: On Educational Inequality, Techno-Solutionism and Inclusive Tools. *Sustainability*, *16*(2), 781. https://doi.org/10.3390/su16020781

Daflizar, D. (2020). Autonomous English Language Learning Beyond the Classroom: Indonesian Tertiary Students' Practices and Constraints. *International Journal of Learning, Teaching and Educational Research*, *19*(10), 194–213. https://doi.org/10.26803/ijlter.19.10.11

Davis, F. D. (1989). Perceived Usefulness, Perceived Ease of Use, and User Acceptance of Information Technology. *MIS Quarterly*, *13*(3), 319. https://doi.org/10.2307/249008

Doshi, A. R., & Hauser, O. P. (2024). Generative AI enhances individual creativity but reduces the collective diversity of novel content. *Science Advances*, *10*(28), eadn5290. https://doi.org/10.1126/sciadv.adn5290

Etkin, H. K., Etkin, K. J., Carter, R. J., & Rolle, C. E. (2025). Differential effects of GPT-based tools on comprehension of standardized passages. *Frontiers in Education*, *10*, 1506752. https://doi.org/10.3389/feduc.2025.1506752

Guzmán-Simón, F., Torres-Gordillo, J. J., & Caballero, K. (2020). Understanding Assessment Processes for Communicative Competence through an Analysis of Teachers' Reported Practice. *Education Sciences*, *10*(4), 116. https://doi.org/10.3390/educsci10040116

Imran, M., & Almusharraf, N. (2024). Google Gemini as a next generation AI educational tool: A review of emerging educational technology. *Smart Learning Environments*, *11*(1), 22. https://doi.org/10.1186/s40561-024-00310-z

Karkoulian, S., Sayegh, N., & Sayegh, N. (2024). ChatGPT Unveiled: Understanding Perceptions of Academic Integrity in Higher Education - A Qualitative Approach. *Journal of Academic Ethics*, *23*(3), 1171–1188. https://doi.org/10.1007/s10805-024-09543-6

Karran, A. J., Charland, P., Martineau, J.-T., de Arana, A. O. de G. L., Lesage, AM., Senecal, S., & Leger, P.-M. (2024). *Multi-stakeholder Perspective on Responsible Artificial Intelligence and Acceptability in Education* (Version 2). arXiv. https://doi.org/10.48550/ARXIV.2402.15027

Kim, J., Yu, S., Detrick, R., & Li, N. (2025). Exploring students' perspectives on Generative AI-assisted academic writing. *Education and Information Technologies*, *30*(1), 1265–1300. https://doi.org/10.1007/s10639-024-12878-7

Knoth, N., Tolzin, A., Janson, A., & Leimeister, J. M. (2024). AI literacy and its implications for prompt engineering strategies. *Computers and Education: Artificial Intelligence*, *6*, 100225. https://doi.org/10.1016/j.caeai.2024.100225

Kong, S.-C., Cheung, M.-Y. W., & Tsang, O. (2024). Developing an artificial intelligence literacy framework: Evaluation of a literacy course for senior secondary students using a project-based learning approach. *Computers and Education: Artificial Intelligence*, *6*, 100214. https://doi.org/10.1016/j.caeai.2024.100214

Leong, J., Pataranutaporn, P., Danry, V., Perteneder, F., Mao, Y., & Maes, P. (2024). Putting Things into Context: Generative AI-Enabled Context Personalization for Vocabulary Learning Improves Learning Motivation. *Proceedings of the CHI Conference on Human Factors in Computing Systems*, 1–15. https://doi.org/10.1145/3613904.3642393

Li, B., Tan, Y. L., Wang, C., & Lowell, V. (2025). Two years of innovation: A systematic review of empirical generative AI research in language learning and teaching. *Computers and Education: Artificial Intelligence*, *9*, 100445. https://doi.org/10.1016/j.caeai.2025.100445

Li, H., Wang, Y., Luo, S., & Huang, C. (2025). The influence of GenAI on the effectiveness of argumentative writing in higher education: Evidence from a quasi-experimental study in China. *Journal of Asian Public Policy*, *18*(2), 405–430. https://doi.org/10.1080/17516234.2024.2363128

Li, Y., Zhou, X., Yin, H., & Chiu, T. K. F. (2025). Design language learning with artificial intelligence (AI) chatbots based on activity theory from a systematic review. *Smart Learning Environments*, *12*(1), 24. https://doi.org/10.1186/s40561-025-00379-0

Liang, J., Huang, F., & Teo, T. (2024). Understanding Chinese University EFL Learners' Perceptions of AI in English Writing: *International Journal of Computer-Assisted Language Learning and Teaching*, *14*(1), 1–16. https://doi.org/10.4018/IJCALLT.358918

Liu, J., Sihes, A. J. B., & Lu, Y. (2025). How do generative artificial intelligence (AI) tools and large language models (LLMs) influence language learners' critical thinking in EFL education? A systematic review. *Smart Learning Environments*, *12*(1), 48. https://doi.org/10.1186/s40561-025-00406-0

Liu, Y., Park, J., & McMinn, S. (2024). Using generative artificial intelligence/ChatGPT for academic communication: Students' perspectives. *International Journal of Applied Linguistics*, *34*(4), 1437–1461. https://doi.org/10.1111/ijal.12574

Lo, C. K., Yu, P. L. H., Xu, S., Ng, D. T. K., & Jong, M. S. (2024). Exploring the application of ChatGPT in ESL/EFL education and related research issues: A systematic review of empirical studies. *Smart Learning Environments*, *11*(1), 50. https://doi.org/10.1186/s40561-024-00342-5

Luat, N. D., Thu, L. P. T., & Thuy, L. T. (2025). Using Gemini for Formative Assessment in English Academic Writing—Critical Insights into The AI Tool's Efficacy. *AsiaCALL Online Journal*, *16*(1), 328–343. https://doi.org/10.54855/acoj.2516117

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

Marpaung, S. (2024). Boosting High School Students' Speaking Proficiency with Gemini AI. *English Journal Antartika*, *2*(1), 8–15. https://doi.org/10.70052/eja.v2i1.543

Menggo, S., Jama, K. B., Adnyani, N. L. P. S., & Krismayani, N. W. (2025). Indonesian English as a Foreign Language Learners' Interpersonal Communication: Profiles and Challenges. *LLT Journal: A Journal on Language and Language Teaching*, *28*(1), 242–256. https://doi.org/10.24071/llt.v28i1.8670

Mingyan, M., Noordin, N., & Razali, A. B. (2025). Improving EFL speaking performance among undergraduate students with an AI-powered mobile app in after-class assignments: An empirical investigation. *Humanities and Social Sciences Communications*, *12*(1), 370. https://doi.org/10.1057/s41599-025-04688-0

Nešić, I., & Hamidovic, K. (2022). Developing a Scale for Assessing Communicative Competence of Students Learning English for Specific Purposes. *Journal of Teaching English for Specific and Academic Purposes*, 481. https://doi.org/10.22190/JTESAP2203481N

Patty, J. (2024). The Use of AI in Language Learning: What You Need To Know. *Jurnal Review Pendidikan Dan Pengajaran*, *7*(1), 642–654. https://doi.org/10.31004/jrpp.v7i1.24609

Patty, J., Lekatompessy, J., & Lekatompessy, F. M. (2025). Implementasi Chatgpt Sebagai Alat Bantu Pengembangan Profesionalisme Guru di SMA Negeri 13 Tiakur. *Jurnal Abdi Insani*, *12*(1), 263–272. https://doi.org/10.29303/abdiinsani.v12i1.2199

Rasheed, A. (2025). Influence of GenAI on Higher Education ESL Students' Writing Confidence. *Qlantic Journal of Social Sciences and Humanities*, *6*(2), 344–351. https://doi.org/10.55737/qjssh.vi-i.25327

Renfeng, J., Gang, Y., & Qi, S. (2025). The Motivational Impact of GenAI Tools in Language Learning: A Quasi-Experiment Study. *International Journal of Applied Linguistics*, *35*(3), 1338–1350. https://doi.org/10.1111/ijal.12701

Sae-Ong, U., & Mohamed Ismail, S. A. M. (2021). Mind Your Gap between Entry English Proficiency and Graduation Expectation. *Asian TESOL Journal*, *1*(1), 1–9. https://doi.org/10.35307/asiantj.v1i1.11

Sayaf, A. M., Alamri, M. M., Alqahtani, M. A., & Alrahmi, W. M. (2022). Factors Influencing University Students' Adoption of Digital Learning Technology in Teaching and Learning. *Sustainability*, *14*(1), 493. https://doi.org/10.3390/su14010493

Shafiee Rad, H. (2025). Reinforcing L2 reading comprehension through artificial intelligence intervention: Refining engagement to foster self-regulated learning. *Smart Learning Environments*, *12*(1), 23. https://doi.org/10.1186/s40561-025-00377-2

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

Song, Y., Huang, L., Zheng, L., Fan, M., & Liu, Z. (2025). Interactions with generative AI chatbots: Unveiling dialogic dynamics, students' perceptions, and practical competencies in creative problem-solving. *International Journal of Educational Technology in Higher Education*, *22*(1), 12. https://doi.org/10.1186/s41239-025-00508-2

Songkram, N., Chootongchai, S., Osuwan, H., Chuppunnarat, Y., & Songkram, N. (2023). Students' adoption towards behavioral intention of digital learning platform. *Education*

and Information Technologies, *28*(9), 11655–11677. https://doi.org/10.1007/s10639-023-11637-4

Tiandem-Adamou, Y. (2025). Using Generative Artificial Intelligence to Support EFL Students' Writing Proficiency in University in China. *Journal of Educational Technology and Innovation*, *6*(4). https://doi.org/10.61414/jeti.v6i4.213

Utami, N. P., Kembaren, F. R., & Daulay, S. H. (2024). Exploring Higher Education Learners' Experience of Utilizing Gemini Chatbot in English Language Learning. *RETORIKA: Jurnal Ilmu Bahasa*, *10*(3), 680–694. https://doi.org/10.55637/jr.10.3.10244.680-694

Wang, C. (2025). Exploring Students' Generative AI-Assisted Writing Processes: Perceptions and Experiences from Native and Nonnative English Speakers. *Technology, Knowledge and Learning*, *30*(3), 1825–1846. https://doi.org/10.1007/s10758-024-09744-3

Wang, C., Zou, B., Du, Y., & Wang, Z. (2024). The impact of different conversational generative AI chatbots on EFL learners: An analysis of willingness to communicate, foreign language speaking anxiety, and self-perceived communicative competence. *System*, *127*, 103533. https://doi.org/10.1016/j.system.2024.103533

Wang, L., Chen, X., Deng, X., Wen, H., You, M., Liu, W., Li, Q., & Li, J. (2024). Prompt engineering in consistency and reliability with the evidence-based guideline for LLMs. *Npj Digital Medicine*, *7*(1), 41. https://doi.org/10.1038/s41746-024-01029-4

Xiao, Y. (2025). The impact of AI-driven speech recognition on EFL listening comprehension, flow experience, and anxiety: A randomized controlled trial. *Humanities and Social Sciences Communications*, *12*(1), 425. https://doi.org/10.1057/s41599-025-04672-8

Yan, L., Greiff, S., Teuber, Z., & Gašević, D. (2024). *Promises and Challenges of Generative Artificial Intelligence for Human Learning* (Version 3). arXiv. https://doi.org/10.48550/ARXIV.2408.12143