Arab World English Journal
INTERNATIONAL PEER REVIEWED JOURNAL ISSN: 2229-9327

Arab World English Journal (AWEJ) Special Issue on CALL Vol. 17, No.12. July 2026 Pp.435-452
DOI: https://dx.doi.org/10.24093/awej/call12.25

AI Literacy as a Predictor of Perceived English Language Development in Generative AI-Supported EFL Contexts

Noof Saleh Alharbi
Language Learning Center, Department of Languages and Translation
Faculty of Arts and Humanities, Taibah University
Madinah, Kingdom of Saudi Arabia (KSA)
Email: nsmalharbi@taibahu.edu.sa

Received: 04/22/2026 Accepted: 07/04/2026 Published:07/15/2026

Abstract
Generative artificial intelligence is increasingly embedded in university students' English language learning, yet its educational value depends on students' AI literacy. This study examined whether first-year EFL undergraduates' AI literacy predicts their perceived English language development in generative AI-supported learning contexts. The study investigated 618 first-year EFL undergraduates enrolled in compulsory English courses at Taibah University, Saudi Arabia. The study examined relationships among AI tool use, AI literacy competencies, and perceived English language development, with attention to differences by gender and academic major. The study is significant because generative AI use in university English courses requires critical, ethical, and purposeful learner engagement. Using a quantitative, cross-sectional predictive correlational survey design, the study assessed AI tool use, AI literacy competencies, and perceived English language development. AI literacy was operationalized across learning-oriented AI use, prompting and task management, critical evaluation and verification, and academic integrity and privacy. Results showed high AI tool use, high AI literacy, and high perceived English language development. No significant differences were found by academic major, and no gender differences appeared in the AI literacy competency total score. However, female students scored significantly higher in critical evaluation and verification. Regression analyses showed that all AI literacy domains significantly predicted perceived English language development, with critical evaluation and verification as the strongest predictors. The total score for the AI literacy competency explained 53.8% of the variance. The study recommends integrating AI literacy instruction into EFL courses, especially verification, responsible prompting, ethical use, and critical evaluation.

Keywords: AI literacy, English as a foreign language, critical evaluation, generative artificial intelligence, perceived English language development

Cite as: Alharbi, N. S..(2026). AI Literacy as a Predictor of Perceived English Language Development in Generative AI-Supported EFL Contexts. *Arab World English Journal (AWEJ) Special Issue on CALL, Vol. 17.* (12).435-452. https://dx.doi.org/10.24093/awej/call12.25

Introduction

Generative artificial intelligence is becoming embedded in higher education because it can generate, revise, translate, summarize, and explain language on demand. In English as a foreign language contexts, this makes it attractive to students who need immediate feedback, vocabulary support, grammar clarification, writing assistance, and opportunities for practice beyond classroom time. Recent studies suggest that AI-mediated learning can support performance, engagement, and motivation when embedded in structured pedagogical designs (Lee et al., 2022; Huang et al., 2023). In EFL learning, AI-supported practice has also been associated with reduced speaking anxiety and improved speaking performance (El Shazly, 2021).

However, the educational value of generative AI cannot be inferred from access or frequency of use alone. UNESCO's guidance on generative AI in education stresses the need for human-centered, ethical, and pedagogically meaningful implementation (UNESCO, 2023). This distinction is important because students may use AI tools frequently without possessing the literacy needed to judge the accuracy, appropriateness, or ethical acceptability of AI-generated output. Much existing research still focuses on students' attitudes, acceptance, or use of AI tools. This leaves a clear research gap: less is known about whether students' AI literacy competencies predict their perceived development in English language learning.

This study is guided by self-regulated learning theory and AI literacy competency frameworks. Self-regulated learning theory explains learning as an active process in which learners plan, monitor, evaluate, and regulate their learning behavior (Zimmerman, 2000; Panadero, 2017). This framework is relevant to generative AI-supported EFL learning because effective AI use requires students to set learning goals, formulate prompts, assess AI output, revise language choices, and decide when AI support is appropriate. In this sense, AI literacy is treated not merely as technical familiarity but as a form of regulated learning behavior.

The study also draws on emerging AI literacy frameworks that conceptualize AI literacy as a multidimensional competence involving understanding, use, evaluation, and ethical judgement (Ng et al., 2021; UNESCO, 2024). UNESCO's AI Competency Framework for Students identifies responsible AI engagement as involving human-centered values, ethics, technical understanding, and applied use (UNESCO, 2024). Accordingly, the present study operationalizes AI literacy across four domains: learning-oriented AI use, prompting and task management, critical evaluation and verification, and academic integrity and privacy.

Saudi Arabia provides a relevant context for this study because higher education is developing within a national environment that foregrounds digital transformation, data, and artificial intelligence as part of Vision 2030. Official national discourse positions data and AI as central to future skills, innovation, and institutional development. In this setting, examining AI literacy among university students is not only a language learning issue but also a higher education readiness issue.

First-year EFL undergraduates are a particularly important group to study. They are transitioning into university academic literacy practices, often come from different disciplinary backgrounds, and are required to complete compulsory English courses that support access to academic study. At this stage, students may be active users of generative AI but may not yet have developed stable strategies for evaluating AI-generated language, detecting errors, protecting academic integrity, or regulating their dependence on automated support. The first year therefore represents a critical point for examining whether AI literacy is associated with students' perceived development in English.

The main aim of this study is to examine whether first-year EFL undergraduates' AI literacy predicts their perceived English language development in generative AI-supported learning contexts. The significance of the study lies in shifting attention from AI access and tool use to AI literacy as a measurable learner competence. This focus is important for EFL curriculum design because students need explicit preparation to use AI critically, ethically, and purposefully rather than relying on AI output uncritically.

The study has five objectives. First, it examines students' reported levels and patterns of AI tool use for English language learning. Second, it assesses students' AI literacy competencies across four domains. Third, it examines students' perceived English language development. Fourth, it tests whether AI literacy competencies differ by gender and academic major. Fifth, it investigates whether AI literacy competencies predict perceived English language development.

The study addressed the following research questions and hypotheses:

1. What are the reported levels and patterns of first-year EFL undergraduates' use of AI tools to support English language learning?
   H1. First-year EFL undergraduates report frequent and varied use of AI tools for English language learning.
2. What are the levels of students' AI literacy competencies across learning-oriented AI use, prompting and task management, critical evaluation and verification, and academic integrity and privacy?
   H2. First-year EFL undergraduates report high levels of AI literacy competencies across the four measured domains.
3. What is the level of students' perceived English language development in generative AI-supported learning contexts?
   H3. First-year EFL undergraduates report high levels of perceived English language development.
4. Are there statistically significant differences in AI literacy competencies by gender and academic major?
   H4a. There are statistically significant differences in AI literacy competencies by gender.
   H4b. There are statistically significant differences in AI literacy competencies by academic major.
5. To what extent do AI literacy competencies predict students' perceived English language development?
   H5. AI literacy competencies significantly and positively predict perceived English language development.

Literature Review

This review argues that the use of generative AI in English language learning should be examined through a competence-based lens rather than narrow indicators such as access, frequency of use, or positive attitudes. It first reviews current research on the educational affordances of generative AI in higher education and language learning. It then highlights the limitations of a research landscape that is still largely shaped by studies of perception, adoption, and productivity. Against this background, the review defines AI literacy for EFL learning as a multidimensional competence encompassing learning-oriented use, prompting ability, critical evaluation, verification, and academic integrity, rather than mere operational familiarity with AI tools. The review therefore establishes the need for a competence-based design to explain how students' AI

literacy relates to their perceived English language development in compulsory first-year EFL courses.

Generative AI in EFL Language Learning

Research on generative artificial intelligence in English language learning increasingly shows that its value depends less on access to tools than on how learners use, evaluate, and regulate AI support. Early and recent work has shown that AI tools can support feedback, drafting, vocabulary development, grammar explanation, and low-stakes language practice. Meniado's (2023) rapid review argues that ChatGPT can support English language teaching, learning, and assessment by providing input, scaffolding output, giving feedback, and supporting fluency development. Similarly, recent meta-analytic evidence suggests that generative AI can positively influence language proficiency and affective-cognitive outcomes, although effects depend on pedagogical design and learner engagement rather than tool availability alone (Saarela et al., 2026).

This distinction is important because AI-supported language learning can either deepen or weaken learning. When students use AI to request explanations, compare alternatives, revise drafts, and reflect on errors, the tool may function as a scaffold. When they use it mainly to generate answers or polished text, the same tool may reduce cognitive effort and obscure actual language development. Recent EFL research, therefore, increasingly treats the use of AI as a matter of learner agency, autonomy, and ethical judgement. Hamamra et al. (2026), for example, found that ChatGPT served as a compensatory resource for EFL writing, but its pedagogical value depended on learners' AI literacy and ethical awareness. This finding directly challenges the assumption that digitally confident students are automatically competent AI users.

From AI Use To AI Literacy

The literature shows a clear movement from studying AI acceptance to studying AI literacy. Perception studies remain useful because they show why students and teachers adopt AI tools. Yet, they do not establish whether learners can use AI in educationally meaningful ways. Recent studies reflect this pattern. Khalid (2026) examined EFL learners' perceptions of AI language tools in a multilingual context, while Sharma and George (2026) investigated teachers' perceptions and experiences of ChatGPT integration in language learning and academia. These studies are valuable for mapping attitudes and concerns, but perception alone cannot explain whether students can prompt effectively, evaluate output, protect authorship, or use AI as a learning resource rather than a substitute for learning.

AI literacy offers a stronger construct because it links tool use to competence. Ng et al. (2021) conceptualize AI literacy as involving understanding, use, evaluation, and ethical awareness. UNESCO's AI Competency Framework for Students similarly frames AI competence as multidimensional, including human-centered values, ethics, technical understanding, and applied use (UNESCO, 2024). In EFL contexts, this means that AI literacy is not simply knowing how to operate a chatbot. It involves using AI to support learning goals, managing prompts and tasks, evaluating linguistic output, and respecting academic integrity and privacy.

Four AI Literacy Domains as Predictors of English Development

The four domains used in the present study are theoretically connected to perceived English language development, as each domain represents a mechanism through which AI use

may be educationally productive. Learning-oriented AI use should predict perceived English development because it directs AI support toward practice, uptake of feedback, revision, vocabulary expansion, grammar awareness, and communicative rehearsal. If students use AI to understand corrections and improve their own performance, AI becomes part of the learning process rather than a shortcut.

Prompting and task management should also predict perceived English development because the quality of AI support depends on the learner's ability to define the task, provide context, ask follow-up questions, and request explanations. Weak prompting may produce generic output, while strategic prompting can generate targeted feedback, examples, reformulations, and practice opportunities. Prompting and task management should also predict perceived English development because learners who can define tasks, request explanations, and ask follow-up questions are more likely to use AI as a scaffold for revision and language learning. This is consistent with recent EFL research showing that the value of ChatGPT-supported writing depends on learners' AI literacy, autonomy, and ethical awareness (Hamamra et al., 2026).

Critical evaluation and verification are especially important in EFL learning because fluent AI output may conceal errors, inappropriate register, weak coherence, or unsuitable rhetorical choices. Moussa et al. (2026) investigated AI-assisted undergraduate EFL writing and showed that AI tools raise questions about authorship, responsibility, and the development of writing ability. Al Hosni (2025) similarly argues that generative AI may affect authorial voice and linguistic authenticity in academic texts. These concerns suggest that students who can critically evaluate AI output are more likely to transform AI assistance into language learning rather than accept polished but unexamined text.

Academic integrity and privacy should predict perceived English development because ethical regulation protects the boundary between assistance and substitution. If students use AI to replace their own thinking, perceived improvement may reflect outsourced production rather than development. If they use AI transparently and responsibly, they are more likely to retain cognitive ownership of learning. This domain is therefore not separate from language development. It shapes whether AI-supported performance can be interpreted as meaningful learning.

Recent Empirical Directions and Remaining Limitations

Recent studies strengthen the need for a competence-based approach. For example, Gümüş and Kara (2025) and related scale-development work show that generative AI literacy can be operationalized as a multidimensional construct rather than treated as a single attitude or use variable. In a recent EFL-focused scale validation study, GenAI literacy is similarly identified as involving knowledge, use, evaluation, design, and ethics (Wang et al., 2025). These studies support the present study's measurement logic, as they show that AI literacy can be empirically assessed across distinct domains.

Recently published research also shows that AI in EFL is no longer a marginal topic. Recent studies have examined learner perceptions, AI-assisted writing, authorial voice, teacher experience, assessment, and systematic reviews of AI in English language classrooms (Al Hosni, 2025; Ennadir & Habiballah, 2025; Khalid, 2026; Moussa et al., 2026; Sharma & George, 2026; Zaidi et al., 2026). However, much of this work remains centered on perceptions, classroom experiences, writing concerns, or assessment practices. Fewer studies test whether specific AI literacy competencies predict students' perceived English language development.

The reviewed literature indicates that generative AI can support English language learning, but its value is conditional on learner competence. Existing studies have examined AI use, attitudes, teacher perceptions, writing support, assessment, and ethical concerns. However, three gaps remain. First, limited research examines AI literacy as a measurable, multidimensional predictor of perceived English language development in EFL contexts. Second, few studies explain why specific AI literacy domains, especially learning-oriented use, prompting and task management, critical evaluation and verification, and academic integrity and privacy, should be associated with perceived language development. Third, there is limited evidence from compulsory first-year university English courses in Saudi Arabia, where students from different majors may use generative AI before receiving systematic training in responsible and learning-oriented AI use. The present study addresses this gap by examining the level of AI tool use, measuring AI literacy across four domains, testing differences by gender and academic major, and investigating whether AI literacy predicts perceived English language development.

Methods

Research Design

This study adopted a quantitative, cross-sectional predictive correlational survey design to examine first-year university students' AI literacy competencies in generative AI-supported English language learning and to test whether these competencies predicted perceived English language development. This design was appropriate because the study measured naturally occurring variables at one point in time and examined statistical associations among them without manipulating any independent variable. Therefore, the study does not make causal claims. Instead, it models whether students' reported AI literacy competencies are statistically associated with, and predictive of, their perceived English language development.

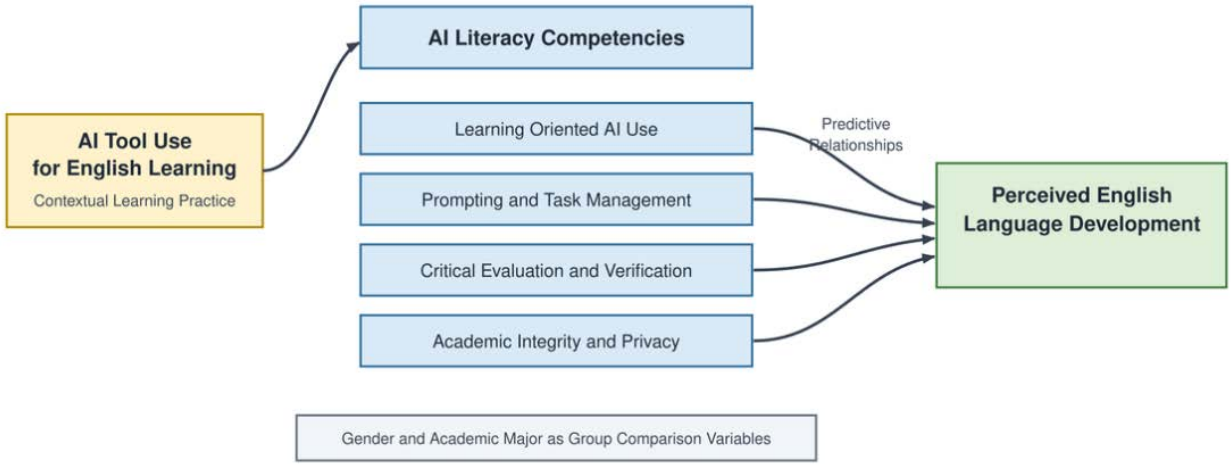


Figure 1. Conceptual Model of the study

Figure 1 presents the conceptual model guiding the study. It positions AI tool use as the learning context, the four AI literacy domains as predictors, and perceived English language development as the outcome. Gender and academic major were included as group comparison variables. The model reflects statistical prediction and association, not causal inference.

Participants

The study was conducted at Taibah University, Madinah, Saudi Arabia. The participants were 618 undergraduate students enrolled in compulsory first-year English courses offered by the English Language Unit in the Department of Languages and Translation, Faculty of Arts and Humanities. Students came from four disciplinary streams, namely Health and Medical Colleges, Scientific Faculties, Humanities Colleges, and Applied Colleges (see Table 1).

Table 1. *Demographic profile of participants*

Variable	Category	n	%
Total sample	First-year EFL undergraduates	618	100
Gender	Male	194	31.4
	Female	424	68.6
Academic major	Health and Medical Colleges	310	50.2
	Scientific Faculties	242	39.2
	Humanities Colleges	45	7.3
	Applied Colleges	21	3.4

The study focused on beginner- to low-intermediate English learners, identified through course placement records and, where available, interpreted using CEFR-aligned descriptors (Council of Europe, 2020). This group was relevant because lower-proficiency learners may benefit from AI support in writing, vocabulary, reading, and speaking practice. Yet, they may also be less able to judge the accuracy, appropriateness, and reliability of AI-generated output.

A non-probability convenience sampling strategy was used because recruitment depended on the availability of first-year English course sections and institutional access. This sampling method is appropriate for educational survey research involving intact student groups, although it limits the generalisability of the findings beyond the study context (Cohen et al., 2011). Students were included if they were enrolled in the target courses during data collection, provided informed consent, and completed the questionnaire with minimal missing data. Responses showing inattentive or patterned answering were excluded.

Research Instruments

Data were collected using three instruments that distinguished between AI exposure, AI literacy, and perceived language-related development. The first instrument was the AI Competencies for English Language Learning Scale (AIC-ELL), which measured four domains: learning-oriented AI use, prompting and task management, critical evaluation and verification, and academic integrity and privacy. The scale was informed by UNESCO's guidance on generative AI and its AI competency framework for students, as well as by multidimensional models of AI literacy that emphasize use, evaluation, and ethical judgement (Ng et al., 2021; UNESCO, 2023, 2024).

The second instrument was the Perceived English Skill Development Scale (P-ESD), which measured students' perceived development in writing, idea organisation, grammar, speaking confidence, oral clarity, language evaluation, and awareness of recurring errors. The third instrument was the AI Use Profile and Controls (AUP-C), which collected background data on AI use frequency, purposes of use, prior training, gender, age, and academic major.

Translation and Piloting

The questionnaire was administered in Arabic; the instrument was prepared in Modern Standard Arabic, with an English version retained for documentation and transparency. Translation

followed a committee-based procedure that included forward translation, reconciliation of wording choices, and cognitive pretesting with a small group of students. This procedure helped ensure that items referring to prompting, verification, academic integrity, privacy, and responsible AI use were clear and conceptually accessible to first-year learners.

Pilot testing was conducted with 20 students who were not included in the main sample. The pilot phase examined item clarity, response distributions, completion time, ambiguous wording, and preliminary reliability. This stage was important because piloting helps identify problems in wording, item interpretation, and response behavior before full administration (Dörnyei, 2007).

Validity and Reliability

Instrument quality was examined through expert review, preliminary evidence of internal structure, and reliability analysis. Face and content validity were established through review by five specialists with relevant expertise in English language learning, educational technology, assessment, and questionnaire design. Their feedback led to the addition of new items, revision or removal of unsuitable items, clearer wording, and stronger alignment between items and their intended domains.

Preliminary evidence of internal structure was examined using the pilot data. Pearson correlation coefficients were calculated between each item and the total score of its corresponding domain, and between each domain and the total scale. The coefficients were statistically significant and positive, suggesting that the items functioned coherently within their intended domains. Given the small pilot sample, these results were treated as preliminary evidence rather than definitive construct validation. This cautious interpretation is consistent with recognized standards for educational and psychological measurement (AERA, APA, & NCME, 2014).

Reliability was assessed using Cronbach's alpha. The AI tool use scale yielded an alpha of .96, the AI literacy competencies scale yielded an alpha of .98, and the perceived English skill development scale yielded an alpha of .95. These coefficients indicate high internal consistency
.

Data Collection Procedures

Data were collected online through a secure survey platform. Students received an information sheet explaining the study's purpose, the voluntary nature of participation, confidentiality procedures, and the right to withdraw without penalty. To reduce careless responding, the questionnaire included one instructed-response item and a minimum completion-time threshold. The study followed established ethical principles for educational research, including informed consent, voluntariness, confidentiality, and the right to withdraw (British Educational Research Association, 2018).

Data Analysis

Data analysis was conducted using SPSS version 23. Descriptive statistics, including means, standard deviations, and item rankings, were used to answer the first three research questions. Pearson item-domain correlations and Cronbach's alpha were used during instrument evaluation to examine preliminary evidence of internal structure and internal consistency.

Because normality tests indicated that the data did not meet the assumptions of normality, non-parametric tests were used for group comparisons. Mann-Whitney U tests examined differences in AI literacy competencies by gender, while Kruskal-Wallis tests examined

differences by academic major. Regression analysis was then used to assess the extent to which AI literacy competencies predicted perceived English language development. Because the study used a predictive correlational design, the regression results were interpreted as evidence of statistical prediction rather than causation.

Results

This section presents the results in relation to the five research questions and the corresponding hypotheses. It begins with the use of AI tools for English language learning, followed by AI literacy competencies and perceived English language development. It then reports group differences by gender and academic major. The final part reports the predictive relationship between AI literacy competencies and perceived English language development.

AI Tool Use for English Language Learning

The first research question examined the reported levels and patterns of first-year EFL undergraduates' use of AI tools to support English language learning. The results showed a high total score for AI tool use (M = 3.82, SD = 0.705), indicating that AI-supported learning was a common feature of students' English learning practices.

Table 2. *AI Tool use for supporting English language learning*

No.	Item	Mean	SD	Level	Rank
1	I use AI tools to learn English or complete English course tasks.	3.97	0.867	High	2
2	I use AI tools to support English writing tasks.	3.96	0.953	High	3
3	I use AI tools to support English reading and comprehension.	3.89	1.075	High	4
4	I use AI tools to support vocabulary learning and usage.	4.09	0.947	High	1
5	I use AI tools to support speaking or conversation practice.	3.87	1.141	High	5
6	I use AI tools to support listening or pronunciation practice.	3.80	1.119	High	6
7	I have received prior training or formal guidance on responsible AI use for study purposes.	3.24	1.289	Moderate	7
	AI tool use total score	3.82	0.705	High	

As shown in Table 2, the highest-rated use of AI was for vocabulary learning and usage (M = 4.09), followed by using AI to learn English or complete English course tasks (M = 3.97), and using AI to support English writing tasks (M = 3.96). Reading and comprehension, speaking or conversation practice, and listening or pronunciation practice were also rated highly. The lowest-rated item concerned prior training or formal guidance on responsible AI use (M = 3.24), which fell within the moderate range.

These results show that students used AI tools frequently and for varied English learning purposes, although formal guidance on responsible AI use was less developed than their actual tool use. Accordingly, H1 was supported. First-year EFL undergraduates reported frequent and varied use of AI tools for English language learning.

AI Literacy Competencies in Generative AI-Supported English Learning

The second research question examined students' AI literacy competencies across four domains: learning-oriented AI use, prompting and task management, critical evaluation and verification, and academic integrity and privacy. The results showed a high total score for AI literacy competency (M = 4.01, SD = 0.652). All four domains were rated in the high range.

Table 3. *AI literacy competencies*

Domain	Mean	SD	Level	Rank
Learning-oriented AI use competency	3.92	0.766	High	4
Prompting and task management competency	4.03	0.779	High	2
Critical evaluation and verification competency	4.01	0.746	High	3
Academic integrity and privacy competency	4.16	0.743	High	1
AI literacy competency total score	4.01	0.652	High	

Table 3 shows that academic integrity and privacy ranked first (M = 4.16), followed by prompting and task management (M = 4.03), critical evaluation and verification (M = 4.01), and learning-oriented AI use (M = 3.92). This pattern indicates that students perceived themselves as competent across all four AI literacy domains, with the strongest self-reported competence in recognising responsible and inappropriate uses of AI.

Accordingly, H2 was supported. First-year EFL undergraduates reported high levels of AI literacy competencies across all four measured domains.

Perceived English Language Skill Development

The third research question examined students' perceived English language development in generative AI-supported learning contexts. The findings showed a high perceived English skill development total score (M = 4.02, SD = 0.734). All items were rated in the high range.

Table 4. *Perceived English language skill development*

No.	Item	Mean	SD	Level	Rank
1	My ability to write coherent English paragraphs improved this semester.	4.03	0.949	High	4
2	My ability to organize ideas and build simple arguments improved.	3.95	1.000	High	7
3	My grammatical accuracy improved after I understood and applied corrections.	4.07	0.904	High	2
4	My confidence in speaking English in short academic situations increased.	3.97	0.968	High	6
5	My ability to produce clearer and more organized oral responses improved.	4.01	0.921	High	5
6	I now evaluate the quality of English sentences instead of accepting them directly.	4.07	0.856	High	1
7	I have become more aware of my recurring language errors and try to correct them.	4.07	0.987	High	3
	Perceived English skill development total score	4.02	0.734	High	

As shown in Table 4, the highest-rated items were evaluating the quality of English sentences instead of accepting them directly (M = 4.07), improving grammatical accuracy after understanding and applying corrections (M = 4.07), and becoming more aware of recurring language errors and trying to correct them (M = 4.07). Students also reported high perceived gains in writing coherent paragraphs, producing clearer oral responses, and confidence in speaking in short academic situations. The lowest-rated item, although still high, concerned organising ideas and building simple arguments (M = 3.95).

These results indicate that students perceived development in both language production and language monitoring. Accordingly, H3 was supported. First-year EFL undergraduates reported high levels of perceived English language development.

Differences in AI Literacy Competencies by Gender and Academic Major

The fourth research question examined whether AI literacy competencies differed by gender and academic major. Before conducting group comparisons, normality was tested using the Kolmogorov-Smirnov and Shapiro-Wilk tests. As shown in Table 5, the normality tests were

statistically significant across gender and academic major categories, indicating that the data were not normally distributed. Non-parametric tests were therefore used.

Table 5. *Tests of normality for AI Literacy by gender and academic major*

Scale	Variable	Category	K-S Sig.	S-W Sig.	Interpretation
AI literacy	Gender	Male	.007	.000	Non-normal
AI literacy	Gender	Female	.004	.000	Non-normal
AI literacy	Major	Health and Medical	.000	.000	Non-normal
AI literacy	Major	Scientific	.000	.000	Non-normal
AI literacy	Major	Humanities	.002	.000	Non-normal
AI literacy	Major	Applied	.002	.003	Non-normal

Gender differences were examined using the Mann-Whitney U test. As shown in Table 6, no statistically significant gender differences were found in learning-oriented AI use ($p = .957$), prompting and task management ($p = .726$), academic integrity and privacy ($p = .106$), or the AI literacy competency total score ($p = .361$). However, a statistically significant gender difference was found in critical evaluation and verification ($p = .025$), with female students showing higher mean ranks than male students.

Table 6. *Mann-whitney U test for differences In AI literacy competencies by gender*

Domain	Male mean rank	Females mean rank	Mann-Whitney U	p value	Result
Learning-oriented AI use	308.93	309.76	41018.000	.957	Not significant
Prompting and task management	305.80	311.19	40411.000	.726	Not significant
Critical evaluation and verification	285.91	320.29	36552.000	.025	Significant, in favor of females
Academic integrity and privacy	292.51	317.27	37832.000	.106	Not significant
AI literacy competency total score	299.81	313.93	39249.000	.361	Not significant

Accordingly, H4a was partially supported. A statistically significant gender difference was found only in critical evaluation and verification, in favor of female students. The hypothesis was not supported for the AI literacy competency total score or the other three domains.

Differences by academic major were examined using the Kruskal-Wallis test. As shown in Table 7, no statistically significant differences were found by academic major in learning-oriented AI use ($p = .161$), prompting and task management ($p = .974$), critical evaluation and verification ($p = .358$), academic integrity and privacy ($p = .229$), or the AI literacy competency total score ($p = .477$).

Table 7. *Kruskal-wallis test for differences In AI literacy competencies by academic major*

Domain	Chi-square	df	p value	Result
Learning-oriented AI use	5.154	3	.161	Not significant
Prompting and task management	0.221	3	.974	Not significant
Critical evaluation and verification	3.228	3	.358	Not significant
Academic integrity and privacy	4.323	3	.229	Not significant
AI literacy competency total score	2.489	3	.477	Not significant

Accordingly, H4b was not supported. AI literacy competencies did not differ significantly by academic major.

Prediction of Perceived English Language Skill Development From AI Literacy Competencies

The fifth research question examined the extent to which AI literacy competencies predicted students' perceived English language development. Regression analysis showed statistically significant predictive relationships for all four AI literacy domains and for the AI literacy competency total score.

Table 8. *Regression coefficients predicting perceived English skill development from AI literacy competencies*

AI literacy domain	r	R^2	F	B	t	Sig.
Learning-oriented use	.601	.360	348.767	.63	10.185	.000
Prompting and task management	.619	.380	382.832	.66	10.129	.000
Critical evaluation and verification	.674	.453	512.818	.69	10.129	.000
Academic integrity and privacy	.646	.416	440.380	.66	11.897	.000
AI literacy competency total score	.734	.538	719.421	.65	14.028	.000

As shown in Table 8, learning-oriented use explained 36.0% of the variance in perceived English language development (R^2 = .360), prompting and task management explained 38.0% (R^2 = .380), critical evaluation and verification explained 45.3% (R^2 = .453), and academic integrity and privacy explained 41.6% (R^2 = .416). Among the four domains, critical evaluation and verification was the strongest domain-level predictor (B = .69). The AI literacy competency total score was also a strong predictor, explaining 53.8% of the variance in perceived English language development (R^2 = .538).

Accordingly, H5 was supported. AI literacy competencies significantly and positively predicted perceived English language development, with critical evaluation and verification emerging as the strongest domain-level predictor. Figure 2 visually summarises the variance explained by each AI literacy predictor, showing that critical evaluation and verification was the strongest domain-level predictor, while the AI literacy competency total score accounted for the largest proportion of variance in perceived English language development.

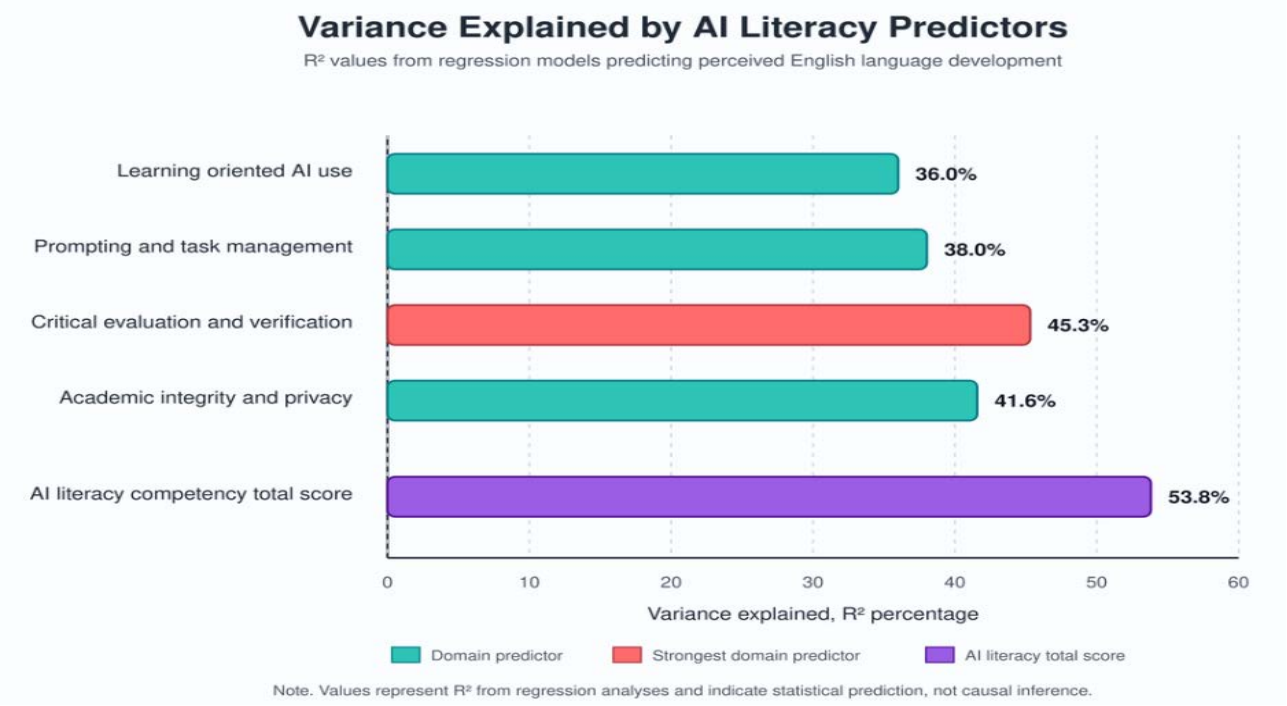


Figure 2. Variance explained by AI literacy predictors

Discussion

RQ1: AI Tool Use in English Language Learning

The first research question examined the reported levels and patterns of first-year EFL undergraduates' use of AI tools for English language learning. The findings showed a high level of AI tool use (M = 3.82, SD = 0.705). Students used AI most frequently for vocabulary learning and usage, followed by English course tasks, writing, reading comprehension, speaking practice, and listening or pronunciation. However, prior training or formal guidance on responsible AI use received the lowest rating and fell within the moderate range. This result indicates that students were active AI users, but their formal preparation for responsible use was weaker than their actual engagement with AI tools.

This finding confirms that generative AI has become a routine resource in students' English learning practices. It is consistent with Meniado's (2023) review, which showed that ChatGPT can support input, output, feedback, and the development of fluency in English language learning. It also aligns with Saarela et al. (2026), who found positive effects of generative AI on language proficiency and affective-cognitive outcomes, although those effects varied by instructional context and learner characteristics. However, frequent use should not be interpreted as well-guided use. The moderate rating for prior training supports UNESCO's (2023) warning that generative AI requires responsible, human-centered, and competence-based implementation. Hamamra et al. (2026) similarly found that ChatGPT can support EFL writing, but its value depends on learners' AI literacy and ethical awareness.

RQ2: AI Literacy Competencies

The second research question examined students' AI literacy competencies across four domains. The findings showed a high AI literacy competency total score (M = 4.01, SD = 0.652). All four domains were rated high. Academic integrity and privacy ranked first, followed by prompting and task management, critical evaluation and verification, and learning-oriented use of AI. This pattern indicates that students perceived themselves as competent in ethical awareness, task management, and evaluation, although learning-oriented use was the least developed domain.

This result supports recent frameworks that conceptualize AI literacy as multidimensional rather than merely operational. Ng et al. (2021) define AI literacy as encompassing understanding, use, evaluation, and ethical awareness, while UNESCO (2024) frames AI competence in terms of human-centered values, ethics, technical understanding, and applied use. Gümüş and Kara's (2025) GenAI literacy scale also supports the view that learning with AI requires more than tool familiarity. The ranking of the domains should still be interpreted cautiously. Academic integrity and privacy may attract socially desirable responses, whereas learning-oriented AI use is more demanding because it concerns whether students use AI to practice, revise, compare, and improve. Wang et al. (2025) similarly treat evaluation and ethics as distinct dimensions of GenAI literacy, supporting the need to teach students how to translate ethical awareness into learning-oriented practice.

RQ3: Perceived English Language Development

The third research question examined students' perceived English language development in generative AI-supported learning contexts. The findings showed a high perceived English skill development total score (M = 4.02, SD = 0.734). The strongest perceived gains concerned evaluating the quality of English sentences, improving grammatical accuracy after applying

corrections, and becoming more aware of recurring language errors. These results suggest that students associated AI-supported learning not only with language production, but also with monitoring, noticing, and self-correction.

This finding is consistent with studies showing that AI can support language learning through feedback, explanation, and practice. Lee et al. (2022) found that an AI-based ChatBot supported performance, self-efficacy, learning attitude, and motivation, while Huang et al. (2023) found that AI-enabled personalized recommendations improved engagement and learning outcomes. In the present study, the strongest perceived gains were linked to sentence evaluation, grammatical accuracy, and error awareness. This suggests that AI may be supporting metalinguistic noticing and self-monitoring. However, the outcome remains perceived development rather than externally measured proficiency growth. The study therefore shows that students believed AI-supported learning contributed to their English development, but it does not establish actual language gains.

RQ4: Differences by Gender and Academic Major

The fourth research question examined differences in AI literacy competencies by gender and academic major. The findings showed no statistically significant gender differences in the AI literacy competency total score, learning-oriented AI use, prompting and task management, or academic integrity and privacy. However, female students scored significantly higher in critical evaluation and verification. No statistically significant differences were found by academic major across the AI literacy competency total score or any of the four domains.

These findings add nuance to the study. The absence of significant differences across academic majors suggests that AI literacy may function as a cross-disciplinary competence in compulsory first-year English courses. This is plausible because all participants were enrolled in shared English courses and faced similar academic language demands, regardless of major. The limited gender difference should not be overstated. Female students' higher score in critical evaluation and verification may indicate stronger evaluative caution or different self-report patterns, but the study design cannot explain the reason. This result should therefore be treated as a basis for future qualitative inquiry rather than evidence of a broad gender effect.

RQ5: AI Literacy as A Predictor of Perceived English Development

The fifth research question examined whether AI literacy competencies predicted perceived English language development. The findings showed that all four AI literacy domains significantly predicted perceived English language development. Critical evaluation and verification was the strongest domain-level predictor ($B = .69$, $R^2 = .453$). The total score for the AI literacy competency explained 53.8% of the variance in perceived English language development. This finding supports the claim that AI literacy is a strong predictor of perceived English development, although the correlational design precludes causal inference.

This is the study's strongest contribution. Prior studies have examined EFL learners' perceptions of AI tools (Khalid, 2026), teachers' experiences of ChatGPT integration (Sharma & George, 2026), AI-assisted writing and authorship (Moussa et al., 2026), and authorial voice in AI-supported academic writing (Al Hosni, 2025). The present study extends this literature by testing whether differentiated AI literacy domains predict perceived English development. The finding that critical evaluation and verification was the strongest predictor indicates that students benefit most when they remain active judges of AI output rather than passive recipients of fluent

text. This supports a scaffold model of AI use rather than a substitution model. In practical terms, EFL courses should teach students how to prompt, verify, revise, and learn through AI while preserving academic integrity and linguistic development.

Conclusion

The main aim of this study was to examine first-year EFL undergraduates' use of generative AI for English language learning, their AI literacy competencies, and the relationship between AI literacy and perceived English language development. The findings provide evidence that generative AI is already firmly embedded in students' English learning practices and that students report high levels of both AI literacy and perceived language development. More importantly, the findings show that AI literacy is not a peripheral variable. It is strongly associated with perceived language development, with critical evaluation and verification emerging as the most influential dimension rather than simple tool use. The main conclusion, therefore, is that educational benefit in generative AI-supported English learning depends on competence, especially evaluative competence, rather than on access alone.

Pedagogical Implications

The pedagogical implications are direct. First, first-year English courses should not treat AI use as an external practice that students manage informally. AI literacy should be taught explicitly within the curriculum. This should include prompting for learning purposes, verification of AI output, revision for voice and appropriateness, and the distinction between assistance and plagiarism. Second, institutions should prioritize formal training because the results showed that training lagged behind actual use. Third, instructional design should focus less on generic encouragement to use AI and more on how to use AI as a scaffold for noticing, practice, reflection, and revision. Fourth, policy should move beyond both prohibition and permissiveness. What is needed is a clear framework for ethical, pedagogically meaningful, and assessment-sensitive use that protects academic integrity while preserving opportunities for authentic learning.

Practical Implications for Instructors and Curriculum Designers

The study also has practical implications for language instructors. Since vocabulary, writing, and grammar-related uses were especially prominent, teachers can build on these existing behaviors by designing guided AI-supported tasks that require students to compare outputs, justify revisions, and explain why one option is more appropriate than another. Because critical evaluation emerged as the strongest predictor of perceived gains, activities should explicitly teach students to question fluency, check contextual fit, and revise AI output rather than merely accept it. This is the point at which language pedagogy and AI literacy should meet. Curriculum designers, similarly, should ensure that AI-related learning outcomes are not limited to digital familiarity but include verification, judgment, and responsible academic use.

Limitations

Some limitations should be acknowledged. First, the study was cross-sectional, so it cannot establish causal relationships. Second, the outcome measure captured perceived development in English skills rather than objective improvement in proficiency. Third, the data were self-reported, which raises the possibility of inflated estimates, especially in ethically sensitive areas such as academic integrity. These issues mean that the findings should be

interpreted as evidence of association and perception rather than as proof of actual language gain attributable to AI literacy.

The study was also conducted within a single institutional context and with a convenience sample, which limits its generalisability. Although the subgroup analyses were informative, the design did not explain why the female advantage in critical evaluation appeared. It also did not include observational or task-based evidence of how students actually used AI during English-learning activities. As a result, the study captures students' reported practices and perceptions more effectively than their real-time behaviors.

Suggestions for Further Research

Future research should address these limitations directly. Longitudinal and experimental studies are needed to determine whether AI literacy leads to measurable gains in writing, speaking, vocabulary, or grammatical accuracy over time. Such studies would help distinguish temporary confidence effects from sustained language development and would make it possible to test whether explicit AI literacy instruction improves actual performance.

Mixed-methods research should examine how students actually use AI in language tasks rather than relying solely on self-reports. Future studies should also compare perceived gains with objective performance measures in order to separate confidence from demonstrated development. In addition, research should explore how proficiency level shapes students' ability to evaluate AI output, since lower-proficiency learners may be especially vulnerable to accepting fluent but inappropriate language. Qualitative work is also needed to examine how students understand the boundaries of academic integrity in practice and how institutional policy shapes those decisions.

Funding: This research received no external funding.

Acknowledgements: Not applicable.

Conflicts of Interest: The authors declare no conflicts of interest.

Authenticity: The authors confirm that this manuscript is an original work.

Artificial Intelligence Statement: During the preparation of this work, the author used Generative AI to support language refinement, organisation, and formatting. The author reviewed and verified all AI-assisted content and retains full responsibility for the manuscript.

About the Author

Noof Saleh Alharbi is an Assistant Professor in the Language Learning Center at the Department of Language and Translation, Faculty of Arts and Humanities, Taibah University, Saudi Arabia. She earned her PhD from the University of Exeter, UK, and her master's degree from Taibah University. Her academic background focuses on English curriculum and teaching methods. Her research interests include ESL/EFL, teaching methods and approaches, and English language skills.ORCID ID: https://orcid.org/0000-0002-4745-9156

References

Al Hosni, J. (2025). Preserving authorial voice in academic texts in the age of generative AI: A thematic literature review. *Arab World English Journal, 16*(3), 244–258. https://doi.org/10.24093/awej/vol16no3.14

American Educational Research Association (AERA), American Psychological Association, & National Council on Measurement in Education. (2014). *Standards for educational and psychological testing*. American Educational Research Association. https://www.testingstandards.net/open-access-files/

British Educational Research Association. (2018). *Ethical guidelines for educational research* (4th ed.). BERA.

Cohen, L., Manion, L., & Morrison, K. (2011). *Research methods in education* (7th ed.). Routledge.

Council of Europe. (2020). *Common European Framework of Reference for Languages: Learning, teaching, assessment: Companion volume*. Council of Europe Publishing. https://book.coe.int/en/education-and-modern-languages/8152-common-european-framework-of-reference-for-languages-learning-teaching-assessment-companion-volume.html

Dörnyei, Z. (2007). *Research methods in applied linguistics: Quantitative, qualitative, and mixed methodologies*. Oxford University Press.

El Shazly, R. (2021). Effects of artificial intelligence on English speaking anxiety and speaking performance: A case study. *Expert Systems, 38*(3), Article e12667. https://doi.org/10.1111/exsy.12667

Ennadir, H., & Habiballah, S. (2025). A comparative study of human and AI-assisted assessment using ChatGPT: The case of Moroccan EFL learners. *Arab World English Journal, 16*(3), 175–187. https://doi.org/10.24093/awej/vol16no3.10

Gümüş, M. M., & Kara, M. (2025). Development and validation of the Generative AI Literacy for Learning Scale (GenAI-LLs). *Australasian Journal of Educational Technology, 41*(4), 1–16. https://doi.org/10.14742/ajet.10236

Hamamra, B., Khlaif, Z. N., & Mahamid, N. (2026). AI literacy and mediated learner autonomy in ChatGPT-supported EFL writing: A qualitative study at a Palestinian university. *International Journal of Educational Technology in Higher Education, 23*, Article 19. https://doi.org/10.1186/s41239-026-00597-7

Huang, A. Y. Q., Lu, O. H. T., & Yang, S. J. H. (2023). Effects of artificial intelligence-enabled personalized recommendations on learners' learning engagement, motivation, and outcomes in a flipped classroom. *Computers & Education, 194*, Article 104684. https://doi.org/10.1016/j.compedu.2022.104684

Khalid, M. W. (2026). Investigating EFL learners' perceptions of the use of AI language tools: A mixed-method study. *Arab World English Journal, Special Issue on Artificial Intelligence, 3*, 195–211. https://doi.org/10.24093/awej/AI3.13

Lee, Y.-F., Hwang, G.-J., & Chen, P.-Y. (2022). Impacts of an AI-based chatbot on college students' after-class review, academic performance, self-efficacy, learning attitude, and motivation. *Educational Technology Research and Development, 70*(5), 1843–1865. https://doi.org/10.1007/s11423-022-10142-8

Meniado, J. C. (2023). The impact of ChatGPT on English language teaching, learning, and assessment: A rapid review of literature. *Arab World English Journal, 14*(4), 3–18. https://doi.org/10.24093/awej/vol14no4.1

Moussa, A., Elkhetabi, S., & Eddahmani, H. (2026). Tracing artificial intelligence in EFL student writing: A linguistic approach to identifying machine-assisted text in undergraduate writing. *Arab World English Journal, Special Issue on Artificial Intelligence, 3*, 73–87. https://doi.org/10.24093/awej/AI3.5

Ng, D. T. K., Leung, J. K. L., Chu, S. K. W., & Qiao, M. S. (2021). Conceptualizing AI literacy: An exploratory review. *Computers and Education: Artificial Intelligence*, *2*, Article 100041. https://doi.org/10.1016/j.caeai.2021.100041

Panadero, E. (2017). A review of self-regulated learning: Six models and four directions for research. *Frontiers in Psychology, 8*, Article 422. https://doi.org/10.3389/fpsyg.2017.00422

Saarela, M., Gunasekara, S., & Kumarage, P. (2026). A meta-analysis of generative AI effects on language proficiency and affective-cognitive outcomes in language learning. *Discover Computing*, *29*, Article 116. https://doi.org/10.1007/s10791-026-10015-1

Sharma, V. K., & George, R. (2026). Teachers' perceptions and experiences of ChatGPT integration in language learning and academia: A mixed-method study. *Arab World English Journal*, *17*(1), 103–115. https://doi.org/10.24093/awej/vol17no1.7

UNESCO. (2023). *Guidance for generative AI in education and research*. UNESCO. https://unesdoc.unesco.org/ark:/48223/pf0000386693

UNESCO. (2024). *AI competency framework for students*. UNESCO. https://unesdoc.unesco.org/ark:/48223/pf0000391105

Wang, Y., Derakhshan, A., & Ghiasvand, F. (2025). EFL teachers' generative artificial intelligence (GenAI) literacy: A scale development and validation study. *System*, *133*, Article 103791. https://doi.org/10.1016/j.system.2025.103791

Zaidi, N. N. A., Hashim, H., & Sulaiman, N. A. (2026). Educators' perceptions and experiences of using artificial intelligence in the English language classroom: A systematic literature review. *Arab World English Journal, Special Issue on Artificial Intelligence, 3*, 180–194. https://doi.org/10.24093/awej/AI3.12

Zimmerman, B. J. (2000). Attaining self-regulation: A social cognitive perspective. In M. Boekaerts, P. R. Pintrich, & M. Zeidner (Eds.), *Handbook of self-regulation* (pp. 13–39). Academic Press. https://doi.org/10.1016/B978-012109890-2/50031-7