



AI-generated self-assessment as a metacognitive scaffold in EFL learning

WAGDI RASHAD ALI BIN-HADY^a  

HAJAR AL SULTAN^b   (Corresponding author)

^aDepartment of English, College of Education, Humanities and Applied Sciences, Hadhramaut University, Yemen; Department of Translation, Faculty of Languages and Translation, University of Aden, Aden, Yemen

^bDepartment of English Language, College of Arts, King Faisal University, Al Ahsa 31982, Saudi Arabia

Abstract

The present study investigates the relationship between English as a Foreign Language (EFL) undergraduates' use of AI-generated self-assessment (AIGSA) and their reported metacognitive skill development. A survey was administered to examine the perceptions of 242 EFL students in Saudi Arabia and Yemen regarding AIGSA and to assess their perceptions of the validity of AIGSA in comparison with instructors' assessments. The quantitative results indicated that EFL students have a high perception of the use of AIGSA in their learning ($M = 3.97$, $SD = .622$). The study also revealed a moderate positive correlation between the perceived use of AIGSA and the reported development of EFL students' metacognitive skills ($r = .432$). The analysis revealed that nationality exerted a significant role in students' perception, where Saudi EFL students reported a higher perception ($M = 4.05$) than their Yemeni counterparts ($M = 3.89$, $Z = 2.166$, $p = .001$). Moreover, the thematic analysis indicated that AIGSA supports learners' self-monitoring and regulation through accessible, detailed, and low-pressure feedback, while its perceived validity remains contextual and complementary to teacher assessment. By foregrounding learners' evaluative knowledge and trust in feedback sources, this study offers novel insights into how AIGSA functions as a metacognitive scaffold rather than a substitute for teacher feedback.

Keywords: AI-generated assessment, EFL learners, metacognitive skills, self-assessment

Introduction

Artificial intelligence (AI) represents the culmination of sustained technological advancement and has increasingly reshaped practices across multiple domains, including education (Ertel, 2024). Within educational contexts, AI has been leveraged to support teacher preparation (Metwally & Bin-Hady, 2025), enhance instructional quality (Alshaie et al., 2025), and transform teaching strategies and assessment practices (Klyshbekova & Abbott, 2024).

Accordingly, much of the existing literature has examined AI from a predominantly teacher-centered perspective, where teachers use AI tools for assessment design, grading, test preparation, and instructional support (Khlaif et al., 2024; Njoku & Wey-Amaewhule, 2025). This focus has been found efficient in assisting teachers to develop assessment-related tasks (Alkhateeb et al., 2025).

In contrast to AI-driven assessment, which typically refers to the use of AI systems by a teacher or institution to evaluate students on summative or formative performance (Sultana et al., 2025), the present study conceptualizes the term “AI-Generated Self-Assessment” (AIGSA), as a learner-centered scaffolding assessment strategy in which learners deliberately use AI tools to evaluate their own language learning performance, interpret the AI assessment, and enhance their performance and learning strategies. AIGSA differs from general AI-mediated feedback, which is featured by passivity or an unsolicited nature (Register, 2026); in contrast, AIGSA is associated with active learners and dynamic use of AI outputs. AIGSA is therefore grounded in metacognitive learning theory (see Zimmerman, 2002), in the sense that the assessment that students receive requires them to monitor their learning, regulate the ways they learn and reflect on their learning experience (Altındağ & Senemoğlu, 2013). As machine learning technologies continue to evolve, scholarly attention has begun to shift toward learner-centered applications of AI that extend beyond traditional teacher-mediated assessment. However, the potential of AI to support learners’ own self-assessment practices—rather than serving solely as a tool for teachers—remains largely underexplored, especially where assessment involves complex cognitive, linguistic, and metacognitive dimensions (Chen et al., 2025). Research indicates that many teachers are insufficiently prepared to systematically guide students in effective self-assessment practices (Chen, 2026; Lee, 2023; Yan et al., 2023). This limitation is particularly concerning given that metacognitive competence and self-assessment (SA) are increasingly regarded as essential for learners to navigate complex, rapidly changing educational and societal contexts (Al Harrasi, 2024; Andrade, 2019; Seema & Padmanabha, 2024). Research shows that EFL learners frequently lack evaluative knowledge and systematic support for self-assessment (Al Harrasi, 2024; Lee, 2023).

As digital learning environments have demonstrated potential in supporting metacognitive skills (Lavrysh et al., 2023), recent advancements in generative AI—such as *ChatGPT*—have further intensified scholarly interest by illustrating AI’s capacity to engage in assessment-related tasks (Bin-Hady et al., 2025). One emerging development is AIGSA, a learner-centered approach that enables students to evaluate their own learning through AI-mediated feedback. Tools like *SelfGauge* harness AI-enhanced project learning to promote students’ self-assessment skills (Zheng et al., 2024), while *Quizizz* enables interactive self-assessment, empowering students to learn English independently (Anggoro & Pratiwi, 2023). Despite its potential, AIGSA has received little attention, possibly due to concerns about teachers’ job security.

As AI tools become increasingly accessible to learners, there is a growing need to understand how students engage with AIGSA, and how such engagement may support metacognitive development. Moreover, learners’ perceptions of the validity of AIGSA in comparison with teacher assessment remain insufficiently understood, particularly across different national contexts.

In response, the present study investigates the use of AIGSA among EFL undergraduates and examines its relationship with metacognitive skill development. In addressing these gaps, we aim to inform the pedagogically sound integration of AIGSA into EFL contexts and contemporary language assessment practices.

Literature Review

Theoretical Framework

This study is explicitly grounded in the intersection of AIGSA and metacognitive theory, positioning AIGSA as the principle independent variable driving the study. While metacognitive skills serve as the overarching dependent variable, the present study hypothesizes that AIGSA may boost EFL students' metacognitive skills by offering iterative, immediate, and individualized feedback (Henderson et al., 2025), which motivates students to regulate, evaluate, and monitor their EFL learning. The link between AIGSA and metacognition is foregrounded in self-regulated learning theory. As Zimmerman (2002) conceptualized, metacognition is an essential component of learning, enabling students to plan, evaluate, and adjust their learning strategies. Therefore, AIGSA functions not only as an assessment/evaluative tool (Lee, 2023), but also as a metacognitive skill tool for enhancing learners' awareness of their own weaknesses, strengths, and the appropriate learning strategies. Nieminen et al. (2025), in their review, concluded that digital assessment enhances learners' self-regulation, digital agency, and identity formation in digital environments. Hooda et al. (2022) added that using AI for assessing students enhances their success. The second variable underpinning our study is metacognitive skills. Metacognition is perceived as "the ability to understand or be aware of one's own or another person's emotions or reasons" (Seema & Padmanabha, 2024, p. 55). Metacognitive skills stand for the abilities that individuals possess to set their own learning goals, design plans for these goals, monitor their progress, and evaluate their performance (Veenman, 2011, 2017). These skills participate in enhancing the 21st century competencies, which are responsible for problem solving, creative thinking, and reaching information (Damayanti et al., 2021). Metacognitive skills play a vital role for learners to monitor, improve, and regulate their own learning through evaluative judgment (Teng & Ma, 2024), learner autonomy and outcomes (Marantika, 2021), and feedback processing (Bernard & Kermarrec, 2022).

However, scholars differed in their classifications of metacognitive skills; they constantly cover core elements such as regulation, knowledge, and experience. Flavell (1979) classified metacognition into four components. *Metacognitive knowledge* is the first skill and refers to the ability of an individual to influence his/her own cognition. *Metacognitive experiences* concern the link between previous events and present occurrences. The third skill is *tasks or goals*, which involves with what the person wants to achieve. The fourth is actions or strategies which refer to the mental or behavioral processes employed to achieve the designed goals.

Moreover, Teng and Ma (2024) confirmed that metacognitive skills have a direct influence on learners to monitor feedback, and they predict their knowledge outcome in academic writing. Additionally, Bernard and Kermarrec (2022) reported that using metacognitive strategies in peer assessment and video feedback leads to the enhancement of the self-regulatory process. Daniar et al. (2023) noted that self and peer assessment are the most effective methods for improving metacognitive skills. In a study of Indonesian students, Marantika (2021) found a significant relationship between metacognitive skills and learning outcomes, observing that students with well-developed metacognitive skills are more likely to enhance their own learning capacity and make deliberate decisions to refine their ideas.

AIGSA as a Student-Centered Feedback and Self-Assessment Tool

A robust body of literature attests on the finding that learners value AI-generated feedback for its accessibility, efficiency, and low emotional risk. A recent finding revealed that students value the feedback of GenAI due to its ease, understandability, timelessness, and lowered risk when compared with teacher-based feedback (Henderson et al., 2025). AIGSA emphasizes learner-centeredness within the learning process, scaffolding learners' ability to use AI tools, evaluate their own performance, and interpret the assessment to enhance their learning strategies.

In the same vein, the existing literature highlights that AI-based assessment can promote student engagement and autonomy. Lee (2023) recruited students to write short essays and then attempt a peer and self-assessment using a rubric provided by their teachers. They then used AI evaluation to score their performance. The results indicated that using AI evaluation contributes to a more sustainable assessment for the learners by facilitating shared evaluation responsibility between teachers and students. Gortaire-Díaz (2026) showed how diverse AI tools, i.e., Twee AI, ChatGPT, and Magic School participate in developing specific pedagogical development in self-directed learning. For example, ChatGPT is found to improve students' writing and grammar, Twee AI is found useful for enhancing their pronunciation and listening, while Magic School is beneficial in enhancing their reading skills.

Assessment has traditionally been viewed as a core instructional tool used by teachers to evaluate learners' progress and inform pedagogical decision-making (Khlaif et al., 2024; Njoku & Wey-Amaewhule, 2025). Beyond its evaluative function, assessment plays a formative role by shaping instructional practices and supporting learners' development through feedback and reflection (Munaroh, 2024). With the increasing emphasis on learner-centered pedagogy, self-assessment has gained prominence as a formative practice that encourages learners to reflect on their performance, identify gaps, and take responsibility for their learning. Self-assessment is closely associated with metacognitive development, particularly the skills of planning, monitoring, and regulation, which are central to self-regulated learning (Ratnayake et al., 2024). Research suggests that metacognitive skills are not acquired instantaneously but develop through sustained practice and reflective engagement (Conway-Smith et al., 2023). However, despite its pedagogical value, effective self-assessment remains challenging in many educational contexts, as learners often lack the evaluative knowledge required to assess their performance meaningfully without adequate scaffolding.

Students' Perceptions of AIGSA in Learning

Recent research has increasingly examined perceptions of AI-generated feedback from both learner and teacher perspectives, revealing a complex and sometimes ambivalent stance toward AI in assessment. Assessment practices have increasingly incorporated AI-powered tools to support teachers in grading, test construction, and performance evaluation (Khlaif et al., 2024; Njoku & Wey-Amaewhule, 2025). These tools are often praised for their efficiency, scalability, and ability to automate routine assessment tasks. For instance, AI applications have been used to assist teachers in generating assessment items and managing large-scale evaluations (Farooq et al., 2024), while validating their ability to measure students' academic needs and emotions (Bhatti et al., 2025). Despite this convenience of AI feedback, Nazaretsky et al. (2026) demonstrate that students systematically perceive human feedback as more credible and higher in quality than AI-generated feedback, even when the feedback content is identical. On the contrary, the study reported that AI feedback strongly supports metacognitive and motivational processes; students seldom acknowledged these strengths. The study highlights that the educational value of AI feedback may be constrained by issues like trust and AI literacy. Additionally, Berg et al. (2025) examined medical students' perceptions and use of AI

feedback and assessment. Students were given a test of eight short-answer questions, and their responses were scored using an AI tool. Using a focus group discussion, the study reported that AI assessment is shaped by a complex interplay between individual, contextual, and technological factors.

The Relationship between AIGSA Use and the Development of Metacognitive Skills

AI tools have also correlated with various variables in existing literature. Alshaie et al. (2025) examined the influence of EFL instructors' AI competence on the quality of English education in Saudi Arabia. The results revealed that EFL instructors' AI competence had a moderate relationship with achieving the quality of English education. Moreover, Alshaie et al. (2025) reported that students' use of AI tools had a moderate relationship with their social and emotional learning ($r = .514$). Moreover, Bhatti et al. (2025) delved into the influence of AI-powered assessment of students' test anxiety and self-efficacy. Their quantitative study recruited 280 students from various universities in Pakistan. The results showed that utilizing AI in the assessment reduces student stress during the test. Moreover, the results reported a positive correlation between using AI-powered assessment and self-efficacy, suggesting that AI boosts student confidence.

Nationality is a core construct in language learning, particularly when technology is integrated. For instance, a study found that Saudi EFL students reported using language learning strategies in a digital environment more frequently than Yemeni learners (Bin-Hady et al., 2020). Notably, most previous research on AI-based assessment has been conducted in Western contexts (Berg et al., 2025; Gortaire-Díaz, 2026; Henderson et al., 2025; Lee, 2023). Consequently, these studies have a contextual gap regarding the examination of the potential correlation between using AI tools in self-assessment and enhancing EFL students' higher-order skills, such as metacognition, in underrepresented contexts such as Saudi and Yemeni EFL learning environments. Our study addresses this gap by examining AIGSA and metacognition in two distinct yet underexplored settings.

Across the current literature, research on AI in educational assessment has largely adopted a teacher-centered perspective, focusing on efficiency, accuracy, and concerns about overreliance on AI tools. Studies in the Saudi context similarly emphasize instructional practices while giving limited attention to learners' emotional, motivational, and cognitive engagement with AI-mediated assessment (Al-Hoorie & AlShakhori, 2025). Although instructors caution that AI may lack contextual and pedagogical judgment (Sahmaniasl, 2025) and remain limited in assessing higher-order skills (Najaf et al., 2025), these concerns primarily reflect teacher perspectives. As AI tools become increasingly accessible to learners, students' perceptions of AIGSA—particularly its validity relative to teacher assessment and its role in supporting metacognitive awareness and learner autonomy—remain underexplored. The present study addresses this gap by shifting the focus from teacher-mediated assessment to examine learners' perceptions of their engagement with AIGSA in EFL learning contexts.

Our study adopts Altındağ and Senemoğlu's (2013) three-dimension model because it operationalizes our method into measurable educational dimensions. The questionnaire items are mapped to each dimension: knowledge (interpreting performance), regulation (enhancing learning strategies), and experience (reflecting on their own learning), which closely correspond with the metacognitive activities used while engaging with AI-based assessment. The model operationalizes the relationship between AIGSA and metacognitive scaffolds. Specifically, AIGSA scaffolds learning among EFL students by making their performance visible (knowledge), promoting their own learning strategies (regulation), and eliciting reflection on their own learning (experience). The model strengthens analysis of the relations between the

two variables by analyzing qualitative data using Braun and Clarke's (2006) six steps to examine how AIGSA scaffolds metacognitive skill development among EFL students and how and when students utilize AIGSA and teacher assessment. Therefore, this model provides a measurable theoretical framework, making it suitable for examining the relationship between EFL students' perception of AIGSA and their reported metacognitive skill development. Specifically, the study seeks to answer the following research questions:

1. To what extent do EFL undergraduates use AIGSA in their learning?
2. To what extent does EFL undergraduates' reported use of AIGSA predict perceived metacognitive skill development?
3. To what extent do Saudi and Yemeni EFL undergraduates differ significantly in their reported perceptions of AIGSA?
4. How do EFL undergraduates perceive the validity of AIGSA as a feedback and self-assessment tool compared to a teacher-provided assessment?

Methods

Research Design

The study employs a convergent correlational mixed-methods design. This is a type of research design that allows simultaneous collecting of quantitative and qualitative data, separate analysis of these, and then comparison of the findings (Creswell & Clark, 2017). The design is also correlational, because it shows the relationship between the EFL students reported use of AIGSA and their perceived metacognitive skill development. A survey instrument combining both quantitative and qualitative items was developed to investigate the perceived relationship between EFL students' AIGSA use and their reported metacognitive skill development. The questionnaire, informed by Voss (2024), assesses EFL students' perceptions of AIGSA, alongside the Metacognitive Skills Scale developed by Altındağ and Senemoğlu's (2013). To complement the quantitative data and capture deeper insights into students' perspectives, open-ended questions were also integrated into the same questionnaire, allowing for an exploration of students' views on both AIGSA and teacher-based assessment in relation to their language learning outcomes.

Participants

The study recruited 242 EFL undergraduate students aged between 19 and 23 from Saudi Arabia and Yemen. The sampling procedures followed a two-stage, country-specific design due to different population size of eligible AIGSA users who responded back to the questionnaire. First, a purposive inclusion criterion (requiring prior AIGSA use) was applied in both settings to identify the eligible population. In Yemen, the entire eligible EFL students using AIGSA and who also completed the questionnaire was included in the sampling frame of the study (a total of 121 students). In Saudi Arabia, where the eligible population who completed the questionnaire was larger, a simple random selection was utilized to draw exactly 121 students, matching Yemen's sample size to enable balanced cross-country comparisons. Thus, every eligible respondent in Yemen was selected, while every eligible Saudi respondent had a non-zero chance of being selected. Specifically, 121 students were selected through a subsequent random selection (see Table 1).

Students reported using AI tools to obtain self-assessment and feedback. ChatGPT was the dominant choice, cited by 90.4% of the participants along with other tools such as Gemini, Twee AI, Copilot, and Poe. They employed direct prompts, such as "correct my mistakes" and "assess my English level." They requested structured support for feedback using prompts like

“in a simple way, create a lesson plan for me.” Students also utilized iterative follow-up prompts, including “explain it again differently” and “provide more examples, until I understand.” To avoid bias or overly positive responses, they explicitly prompt the AI tools to “be frank and not just praise me.”

All participants were enrolled in the Bachelor of Arts, Education and Translation programs at various universities in both Saudi Arabia and Yemen and were required to have used AIGSA in their language learning. Because all Yemen EFL respondents were included, they represent the full target population of AIGSA using EFL undergraduates in Yemen, whereas the Saudi subsample represents a randomly drawn subset of a larger eligible population. They attended various universities in Saudi Arabia and Yemen. The sample was predominantly female, comprising 87.6% of the participants, while male students accounted for only 12.4% (see Table 1). Ethical approval for the study was granted by the Deanship of Scientific Research at King Faisal Univeristy. Students provided their consent by ticking their agreement on a statement to participate in the study. They chose to consent to their involvement in this study.

Table 1. *Participants’ background information and AI engagement*

Categories		Frequency	Percentage %
Informed consent	I agree	242	100.0%
	I don’t agree	0	0.0%
	Total	242	100.0%
1. What is your country?	Saudi Arabia	121	50.0%
	Yemen	121	50.0%
	Total	242	100.0%
2. What is your gender?	Male	30	12.4%
	Female	212	87.6%
	Total	242	100.0%
3. Do you use AI in your English language learning?	Yes	242	100%
	No	0	0%
	Total	242	100.0%

Instruments

To address the research questions, we have designed a closed-ended questionnaire with two open-ended questions, based on previous studies. The first part of the questionnaire gathers demographic information, including major, field of study, and university. The second part aimed to collect EFL undergraduate students’ perceptions regarding the use of AI in assessing their language learning. The researchers reviewed Slimi et al. (2025) and Hooda et al. (2022) and extracted key themes related to AI-based assessment. Guided by these themes, the researchers coined a set of seven items tailored to the present context and study’s purpose (see Table 5).

The third part focused on collecting EFL students’ self-reported perceptions of their meta-cognitive skill development. It comprised thirty items categorized into three subsections taken from Altındağ and Senemoğlu’s (2013) Metacognitive Skills Scale in a second language: metacognitive knowledge, metacognitive regulation, and metacognitive experiences. All the

questionnaire items were translated into Arabic to ensure clarity for the participants. The questionnaire employs a five-point Likert scale ranging from “strongly disagree” to “strongly agree.” The final section of the questionnaire included two open-ended questions designed to address the research question concerning the perceived validity of AIGSA in comparison with teacher assessment.

Specifically, perceived validity was explored through students’ reported perceptions of the assessment source and their perceived level of learning support provided by AI- and teacher-generated feedback. The qualitative analysis of these two questions in the questionnaire is conceptually consistent with the adopted model, enabling the study to examine how AIGSA functions as a metacognitive scaffold that supports regulation and experience, while contributing to the gradual development of metacognitive knowledge.

To ensure the content validity of the questionnaire, the researchers submitted the final copy of the questionnaire to two university professors in AI-technology and Applied Linguistics using predefined criteria on items’ clarity, relevance, appropriateness, and construct coverage. They provided detailed feedback on some items, suggesting revision for ambiguous wording. They suggested adding two open-ended questions to the questionnaire. We adjusted the questionnaire before the final administration.

After adjusting the questionnaire items based on expert feedback, the researchers conducted a pilot study with 50 students to assess item clarity and appropriateness to the target sample. The results showed students understood the items without any difficulty, as no requests for clarification were reported during pilot administration. These participants were later excluded from the final analysis. To establish construct validity, we performed item-dimension correlation analysis using Spearman correlation coefficients. The results demonstrate that all items in the four factors exceeded the recommended .30 threshold. Moreover, the mean correlation for the categories ranged from .472 to .678, suggesting acceptable to strong convergence of the items with respect to their respective constructs. These results provide evidence that the adopted scale is sufficiently valid to measure AIGSA and metacognitive skills development among EFL students (see Table 2 and Appendix A).

We used Cronbach’s alpha to measure the internal consistency of the items after the pilot study. The overall questionnaire yields a Cronbach’s alpha of .89, indicating good consistency. As the questionnaire comprises two factors, the first category (EFL students’ AIGSA use) achieved a reliability level of .79, indicating acceptable consistency, while the second category (metacognitive skills) yielded a reliability level of .83, with subscale internal consistencies ranging from acceptable to good: metacognitive knowledge ($\alpha = .81$), metacognitive regulation ($\alpha = .73$), and metacognitive experience ($\alpha = .67$) (see Table 2).

Table 2. *Construct validity and reliability scores*

Variables	Subdimension/ Component	Cronbach’s α	Items	Correlation range	Main correlation
AIGSA	— (standalone)	.79	1-7	.440 – .764	.678
Metacognitive skills	Overall Metacognitive Skills	.83	8-36		
	└ Metacognitive knowledge	.81	8-15	.531 – .768	.676
	└ Metacognitive regulation	.73	16-27	.415 – .574	.499
	└ Metacognitive experience	.67	28–36	.375 – .553	.472
Overall		.89			

Data Analysis

Quantitative data were analyzed using SPSS (Version 30). Descriptive analysis was used to calculate mean scores and standard deviations. Inferential analysis was performed using linear regression to examine the perceived influence of AIGSA on metacognitive skills as perceived by EFL students. Moreover, we first examined the data distribution using the Shapiro-Wilk test ($W = .96, df = 242, p < .001$). The results indicated a violation of the normality assumption. Consequently, a non-parametric Mann-Whitney U test was utilized to determine whether nationality significantly influenced students' perceptions of using AIGSA. The Mann-Whitney U test is suitable when the data distribution assumption between the two independent groups is not met (Nachar, 2008). Effect size is also reported. Table 3 details the analysis plan, linking each research question to its variable, statistical test, and interpretation criteria.

Table 3. Analysis criteria of quantitative and qualitative data

Research question	Variable	Statistics	Interpretation criteria
RQ1	AIGSA perception score (7 items)	Descriptive statistics	Mean ≥ 3.5 = high; 2.5–3.4 = moderate; < 2.5 = low
RQ2	Prediction of AIGSA to metacognitive skill development (29 items)	Linear regression	$p < .05$; small ($r = .10-.29$), medium ($r = .30-.49$), and large ($r \geq .50$)
RQ3	Nationality (Saudi/Yemeni); AIGSA perception score	Mann-Whitney U test	$p < .05$ = significant difference; effect size (r) reported
RQ4	Validity of AIGSA compared with teacher-provided assessment	Thematic analysis	Major themes identified; comparison of positive/neutral/negative perceptions

Qualitative data were obtained from two open-ended questionnaire items examining (a) students' trust in AI-generated versus teacher assessment and (b) the type of feedback that made them feel most supported in learning. Responses were analyzed manually using reflexive thematic analysis following Braun and Clarke's (2006, 2019, 2020) six-phase framework. Each phase of the analysis is transparently described in detail. Phase 1–Familiarization: we read and reread all the responses multiple times to familiarize ourselves with the nature of the responses. We found that responses ranged from single-word answers (e.g., “Teacher,” “AI,” “Both”) to extended reflective explanations. During this phase, attention was paid to the words they used. An inductive, data-driven approach was used.

Phase 2–Inductive first-cycle inductive coding was conducted to capture participants' meanings, resulting in approximately 54 initial codes across both questions. Coding focused on identifying patterns related to trust, perceived validity, emotional and cognitive support, and learners' regulation of feedback sources. Examples of initial codes included: “AI explains in exhaustive details,” “AI explains my mistakes in a simple way and clarifies the correct steps,” “AI is good for practice, but the teacher is better for real evaluation,” “They complete and support each other,” “The teacher gives feedback at school... while AI is for when I'm at home,” and “Both, depends on my needs.” Later, we combined related codes into broader categories.

Phase 3–Grouping codes into categories. In this phase, we generated initial themes by categorizing related codes into broader categories. For example, “AI explains my mistakes in a simple way and clarifies the correct steps” and “AI gives me more than one source for learning,” were categorized under the subtheme “quality & depth of explanation.” Likewise, the codes “It explains my mistakes in a simple way and clarifies the correct steps,” “AI gives it to me in a way that my brain understands,” were attached under the category “personalization & adaptability metacognitive knowledge.” Nearly six subthemes were identified at this stage (see Tables 9 and 10). Phase 4–Reviewing and refining themes: categories (subthemes) were iteratively reviewed, refined, and grouped into two broader themes. The first theme, “validity of feedback,” includes broader categories like objectivity and fairness, personal knowledge and context, quality and depth of feedback, or emotional support. This theme directly corresponded to the research questions. The second major theme, “AIGSA as a metacognitive tool” comprises categories: personalization & adaptability, metacognitive knowledge, and human-AI complementarity & contextual use, which corresponded with the metacognitive learning strategies. Phase 5–Defining and naming themes: In this phase, each major theme was checked against its subthemes. We provide direct audit trails in Table 8 (below) beginning with raw data (initial responses), two categories, and major themes. We also provide the frequencies of each code (number of mentions). Phase 6–Writing the report: we finalize the analysis process by presenting illustrative quotes in Table 9 and Table 10 below.

To ensure credibility and trustworthiness in the analysis, we employed some procedures. The first and second researchers conducted the analysis independently. The first researcher generated 60 initial codes, whereas the second generated 52. Comparison of the coding outputs revealed 45 codes yielded an intercoder agreement of approximately 80.4%. This independent coding was followed by a discussion session engaging both researchers to resolve disagreements on some codes and theme development. Each disagreement was solved by referring to the raw data, ensuring that not a single code was missing in our coding until consensus was reached. This discussion session resulted in a final code framework of approximately 54 agreed-upon codes. We then independently grouped the codes into categories and subthemes, and subsequently developed and named the major themes. We separately carried out this process to minimize bias and enhance analytical rigor. These procedures were systematically recorded in an audit log, ensuring traceability of all analytical decisions. The analysis was conducted bilingually (Arabic and English), with meanings preserved through careful interpretation. Research triangulation and the audit trail align with the recommendation practices for trustworthiness used in reflective thematic analysis (Braun & Clarke, 2020). The findings are illustrated in Table 4 below.

Table 4. Themes and sub-themes emerged in students’ responses

Major themes	Sub-themes
1. Validity of feedback sources	1. Psychological safety & accessibility 2. Contextual and situational trust and support 3. Quality & depth of explanation 4. Practical efficiency 5. Emotional & Motivational Support
2. AIGSA as a metacognitive tool	1. Personalization & Adaptability 2. Metacognitive knowledge 3. Human-AI Complementarity & Contextual Use

Findings

RQ1. To What Extent do EFL Undergraduates use AIGSA in their Learning?

Table 5 demonstrates the Yemeni and Saudi EFL students’ use of AI tools for obtaining self-assessment and feedback on their learning. The participants reported a high level of use with an overall mean score of 3.97 and a standard deviation of .622. The highest-rated item indicated that participants perceived AI-powered assessment as helping in reducing their anxiety and stress when evaluating their English language skills ($M = 4.12$, $SD = .857$). The standard deviation seems to be higher than the mean average SD (.622), which indicates a moderate variability of students’ responses on this item. In contrast to other indicators, students reported the lowest mean perceptual score of 3.69 for using AI tools to track their progress. Notably, though this was the lowest mean score documented, it is still interpreted as a relatively high mean score based on the scale.

Table 5. EFL students’ use of AIGSA in their learning

Statements	Mean	SD
1. Interactive and engaging AI assessments increase my motivation to learn English.	4.01	.890
2. AI-based self-assessment enhances my engagement and participation in English language learning activities.	3.97	.908
3. Using AI-generated assessment encourages me to take control of my own learning process.	4.11	.875
4. Using AI-generated feedback help me adjust my English learning strategies and improve performance.	3.97	.926
5. Using AI-generated assessment enhances my achievement and grades.	3.89	.931
6. AI helps me track my progress and set realistic goals for learning English.	3.69	1.02
7. AI-powered assessments have reduced my anxiety and stress when evaluating my English language skills.	4.12	.857
Overall	3.97	.622

RQ2: To What Extent does EFL Undergraduates’ Reported use of AIGSA Predict Perceived Metacognitive Skill Development?

The linear regression analysis presented in Table 6 revealed a moderate association between EFL students’ reported use of AI-generated self-assessment and their perceived metacognitive skill development ($\beta = 0.432$, $t = 7.416$, $p < .001$). The model, i.e., EFL students’ reported use of AIGSA, explained only 18.6 % of the variance in their metacognitive skills ($R^2 = 0.186$, $F = 54.99$, $p < .001$), indicating that EFL students use many other factors which may contribute to development of their metacognitive skills. Furthermore, the unstandardized coefficient ($B = 0.279$) suggests that for every one-unit increase in students’ use of AIGSA, their metacognitive skill scores increased by 0.279 units. However, the data provided is just perceptual and does not address the real impact of applying AIGSA on their metacognitive skill development (see Table 6).

Table 6. *The relationship between the use of AIGSA and EFL students’ metacognitive skill development*

Model	Unstandardized coefficients		Standardized coefficients	T	Sig.	R	R square	F	Sig.
	B	Std. error	Beta						
(Constant)	2.755	.151	.432 ^a	18.24	.000	.432	.186	54.99	.000 ^b
(AIGSA)	.279	.038							

a. Dependent Variable: metacognitive skills.
b. Predictors: (Constant), AIGSA.

RQ3. To What Extent do Saudi and Yemeni EFL Undergraduates Differ Significantly in their Reported Perceptions of AIGSA?

A total of 242 EFL students (Saudi = 121, Yemeni = 121) participated in the study. Table 5 shows that Saudi EFL students reported a slightly higher perception with a mean score of 4.05 (*SD* = 0.659) than the Yemeni EFL students, who recorded a mean of 3.89 (*SD* = 0.574). As the data did not follow the normal distribution, we used the Mann-Whitney U test to check whether the difference was statistically significant or not. The Asymp Sig. (2-tailed) value in Table 5 shows that the difference was statistically significant ($Z = 2.166$, $df = 240$, $p = .030$). Due to the small difference between Saudi and Yemeni students ($M = 4.05$ vs. 3.89), it is helpful to report and interpret the effect size using Cohen’s *d*. The result yields 0.27, suggesting a small predictive effect. The confidence interval (95%) for the mean difference, as seen in Table 7, ranged from .01102 to .32429, indicating that EFL students’ nationality exerts a small influence on their perceptions toward using AI-generated content in their self-assessment (see Table 7).

Table 7. *The relationship between students’ nationality and use of AIGSA*

Nationality	N	Mean	SD	Z Value	Asymp Sig. (2-tailed)	95% Confidence interval of the difference	
Saudi	121	4.05	.659	2.166	.030	Lower	Upper
Yemeni	121	3.89	.574			.01102	.32429

RQ4. How do EFL Undergraduates Perceive the Validity of AIGSA as a Feedback and Self-assessment Tool Compared to a Teacher-provided Assessment?

Out of the 242 respondents, 129 ¹students expressed primary trust in teacher assessment. In contrast, 50 students ²reported greater trust in AI-generated assessment. Notably, 34 participants articulated conditional trust, indicating that their preference depended on contextual factors such as task type, learning purpose, or assessment stakes. In addition, 29 students reported trusting both AI-generated and teacher assessment, reflecting a blended approach to feedback use (see Table 8).

¹ 43 simply stated “teacher” without explanation.
² 19 out of 50 students gave a single-word response (“AI”) without justification.

Table 8. *Frequencies of respondents to the kind of assessment they preferred*

Responses	Teachers	AIGSA	Conditional trust	Both AIGSA & teachers	Total
Frequencies	129	50	34	29	242

The thematic analysis³ also identified two major themes. The first theme, validity of feedback sources, directly corresponds to the research question and the open-ended questions, which explicitly elicited students’ evaluations of trust and perceived support from AI-generated and teacher feedback. The second theme, AIGSA as a metacognitive tool, which concerns students’ engagement in metacognitive skills, emerged inductively from the data. Although not explicitly targeted by the open-ended questions, this theme was evident in students’ explanations and justifications. Below, each theme is explained further with illustrative quotes.

Validity of Feedback Sources

Although EFL undergraduates perceive AIGSA as useful and credible, the qualitative findings indicate that its validity is contextual and situational. While the thematic analysis highlights that teachers continue to be perceived as the most valid and supportive source of assessment, it also shows how students do not conceptualize validity as exclusive but rather adopt a strategic and situational approach to assessment. AIGSA is increasingly trusted as a complementary and context-dependent assessment tool. EFL undergraduates’ engagement with AIGSA was primarily influenced by several factors, including psychological safety and accessibility (40 mentions), followed by quality of explanation (34), contextual trust (34), practical efficiency (22), and emotional and motivational support (20). Together, they indicate that AIGSA is valued as a low-pressure, supportive, and cognitively effective tool rather than a substitute for teacher judgment. Table 9 below presents these factors that were identified in our data as sub-themes, along with illustrative quotes (see Table 9).

Table 9. *Sub-themes of EFL students’ preference of assessments*

Subthemes	Initial responses and frequencies
Quality & Depth of Explanation	1. AI explains in exhaustive details (15 times). 2. AI explains my mistakes in a simple way and clarifies the correct steps (11 times). 3. AI gives me more than one source for learning (8 times).
Contextual and situational trust	1. It depends on the teacher (8 times). 2. If it’s an exam, I trust the teacher more (6 times). 3. AI is good for practice, but the teacher is better for real evaluation (14 times). 4. Sometimes AI, sometimes the teacher (1 time). 5. It depends on the task (5 times).
Practical Efficiency	1. AI gives me the answer directly (8 times). 2. AI gives me a summary of the slide or page (4 times). 3. AI helps me with many things and makes tasks easier (10 times).
Emotional & Motivational Support	1. AI encourages a lot and makes you feel you are the best (14 times). 2. AI tells the truth and does not side with another party (6 times).

³ The TA and the emerging themes and sub-themes were based on the explanations and justifications students provided for both open-ended questions.

Students viewed trust and support as situated, negotiated, and contextual, evaluating the validity of AIGSA and teacher assessment according to task demands, learning purpose, and assessment stakes. The illustrative quotes suggest that students perceive AIGSA as valid in contexts requiring immediate clarification and self-assessment. Students associate feeling supported with empathy, motivation, and psychological safety, not just information. AIGSA was viewed as effective and supportive for grammar checking, explanation, and self-assessment, while teacher assessment was reserved for more holistic and performance-based evaluation. This distinction was frequently articulated through expressions of conditional trust, such as “it depends,” “sometimes AI, sometimes the teacher,” and “AI for practice, teacher for evaluation.” These responses indicate that students’ judgments of assessment validity were situational and task-dependent.

Students also explicitly differentiated between skill types, perceiving teachers’ assessment as more valid for speaking, pronunciation, real classroom performance and interaction, and holistic evaluation, where contextual understanding and observation over time were considered essential. These are illustrative quotes:

“The teacher knows my real level.” [Student 105]

“I prefer the teacher, because through him I discover my pronunciation and grammatical mistakes.” [Student 138]

“I prefer face-to-face interaction makes the assessment more accurate.” [Student 38]

“I prefer the teacher because he sees me in class and knows how my English really is.” [Student 7]

“I would trust my teacher more when assessing my English skills, because teachers can understand my personality, progress, and effort.” [Student 89]

These quotes indicate that students associate the validity of teachers’ assessment with teachers’ ability to provide personalized, context-sensitive feedback based on sustained interaction with the learner.

AIGSA as a Metacognitive Tool

While the quantitative results confirm that AIGSA is significantly associated with metacognitive skill development, the qualitative findings demonstrate that AIGSA appears to function as a metacognitive monitoring tool. As shown in Table 10, there the two subthemes: personalization and adaptability and human-AI complementarity and contextual use highlight students’ active engagement in monitoring, evaluation, and regulation.

Students’ emphasis on AI’s detailed explanations, immediate feedback, and repeated questioning reflects active engagement in metacognitive monitoring. The illustrative quotes in Table 10 show that AIGSA supports metacognitive knowledge, regulation, and experience by enabling students to monitor understanding, evaluate errors, regulate strategies, and strategically integrate feedback sources. AI is experienced as cognitively supportive, especially for self-study and clarification, even if emotionally neutral. Students reported using AI to identify errors, test understanding repeatedly, compare answers, and regulate pace and strategies. Participants explained such processes succinctly:

“I can repeat asking AI questions until I understand everything.” [Student 10]

“I prefer AI, because I can choose the best way of learning for me. I don’t have to adapt to a teacher’s own way.” [Student 80]

“It explains my mistakes in a simple way and clarifies the correct steps.” [Student 47]

Table 10. *EFL students’ perceptions of AIGSA as a metacognitive tool*

Theme	Sub-themes	Initial responses and frequencies
AIGSA as a Metacognitive Tool	Personalization & Adaptability	1. It explains my mistakes in a simple way and clarifies the correct steps (11 times).
	Metacognitive knowledge	2. AI explains in exhaustive detail (15 times). 3. AI gives it to me in a way that my brain understands (9 times). 4. AI gives me more than one source for learning (8 times). 5. Because I can choose the best way of learning for me, I don’t have to adapt to the teacher’s own way (7 times).
	Human-AI Complementarity & Contextual Use	1. They complete and support each other (18 times). 2. The teacher gives feedback at school... while AI is for when I’m at home) (8 times). 3. Both, it depends on my needs (8 times). 4. Sometimes the teacher and sometimes the AI (5 times).

Such statements reflect iterative cycles of monitoring and evaluation, core components of metacognitive regulation. Together, these behaviors correspond directly to core metacognitive processes, even though students did not explicitly label them as such. This positions AIGSA as a metacognitive scaffold for independent learning rather than a substitute for teacher assessment.

Discussion

The study aimed to examine the extent to which EFL students perceive using AIGSA in their English language learning. The results indicate that Saudi and Yemeni EFL students reported a high level of AI tool usage for self-assessment and feedback ($M = 3.97$). This finding suggests that EFL students are highly aware of the potential of AI tools to support autonomous and life-long learning. Such a high mean score may demonstrate the accessibility of AI-driven platforms that offer immediate feedback and assessment opportunities for students’ metacognitive skill development. While broadly consistent with previous research findings on the positive learner perceptions of AI-supported learning environments, Lee (2023) found that AI-based self-assessment helps facilitate the evaluative responsibility between teacher and students. Similarly, Gortaire-Díaz (2026) pointed out that AI tools improve students’ language skills, such as writing and listening, as well as language systems such as grammar and pronunciation. Likewise, Berg et al. (2025) reported that AI-based assessment is useful and more reliable. However, our findings provide empirical evidence from under-represented contexts, Yemen and Saudi Arabia, where digital infrastructure differs constantly. Moreover, prior studies reported general use of AI tools (Bin-Hady et al., 2024), or teacher-focused AI technology competence (Alshaie et al., 2025). The current study examines the relationship between AI-generated self-assessment and metacognitive skills; a connection rarely explored in previous studies. The findings therefore suggest that perceived AIGSA usefulness in EFL contexts extends beyond their focus to encompass metacognitive skill development.

In response to the second research question, the study revealed a moderate positive association between the reported level of AIGSA use among Yemeni and Saudi EFL students and their perceived metacognitive skills ($r = .432$). Beyond indicating a statistical relationship, this finding suggests that AIGSA may function as a process-oriented learning environment

in which metacognitive engagement is actively enacted rather than passively supported. Therefore, the value of this finding lies not only in mirroring previous findings such as Bhatti et al. (2025), who reported that AI-powered assessment contributes to reduced anxiety and enhanced self-efficacy, but also in illustrating how such benefits enhance their interaction and develop their metacognitive regulation. Interestingly, the qualitative findings extend the line of research by demonstrating the mechanism by which AIGSA may foster metacognitive awareness, monitor understanding, identify errors, and strategically regulate feedback use. From a metacognitive perspective, students repeatedly describe using AIGSA to check explanations, identify errors without fear, and adjust their learning strategies. This finding suggests that AIGSA operates as a self-regulatory feedback loop. This explains Altındağ and Senemoğlu's (2013) ideas that metacognitive development involves not only knowledge of strategies but also regulation and experiential awareness. A key contribution of this finding is the conceptual positioning of AIGSA as a construct that is not only correlated with metacognitive skills, but also instantiates metacognition activity through engagement, monitoring, and self-regulation supported by immediate and repeatable AI feedback. This proposes AIGSA as a pedagogical technique for developing self-regulating learning in EFL in various contexts.

Additionally, to answer the third research question, the study observed a statistically significant but small nationality-based difference in EFL students' reported use of AIGSA, with Saudi EFL students reporting higher usage than Yemeni students. However, the effect size was small, which limits the practical magnitude of this difference. Such a finding demonstrates a modest tendency for Saudi students to report greater engagement with AI technology in their learning over than Yemeni students. This finding aligns with prior research on technological infrastructure and learning strategy use (Bin-Hady et al., 2020). Saudi students' higher perceptions of AIGSA may reflect greater exposure to digital learning environments, which in turn supports more frequent engagement in AI-mediated metacognitive practices. This contextual factor reinforces Berg et al.'s (2025) conclusion that AI assessment use is shaped by a complex interaction of individual, contextual, and technological variables.

In responding to the fourth question, students' contextual evaluation of AIGSA validity helps explain why high perceptions of AI use coexist with continued reliance on teacher assessment. The present findings both confirm and extend this perspective. While students in this study clearly continued to express greater trust and perceived support in teacher assessment, particularly for contextualized and performance-based assessment, they also reported using AIGSA as a supportive tool for clarification, practice, and independent monitoring. Students often find AI supportive, but do not always consider it valid for final or holistic assessment. Our findings extend previous research (Al Harrasi, 2024; Andrade, 2019; Nazaretsky et al., 2026; Veenman, 2017) by positioning AIGSA not as a replacement for teachers. Students' conditional trust in AI and strategic selection between AI and teacher feedback indicates active metacognitive control, a key indicator of advanced self-regulated learning (Veenman, 2017). In contrast to previous studies (e.g., Al Harrasi, 2024; Andrade, 2019; Nazaretsky et al., 2026), which reported that students struggle with self-assessment when evaluative knowledge is underdeveloped or teacher support is either insufficient or excessive, our findings revealed that EFL learners' evaluations of AIGSA incorporate emotional, relational, and affective dimensions of feedback. This contribution suggests that the perceived usefulness of AIGSA is multidimensional, with learners using AI as a metacognitive tool to enact valid perceptions.

While other works (Lee, 2023; Najaf et al., 2025), emphasize teacher guidance as the primary mechanism for developing evaluative knowledge, we find that AIGSA may serve as a digital scaffold that supports evaluative knowledge, self-efficacy, and autonomous engagement in

self-assessment even beyond formal classroom mediation. Although prior studies stressed AI's efficiency (Lee, 2023; Najaf et al., 2025), our qualitative data reveal that students perceive AIGSA's validity as task-dependent. Students valued AIGSA for grammar, clarification, and practice-oriented tasks, but teacher assessment preferred performance-based and communicative skills. This supports Sahmaniasl's (2025) concern that full reliance on AI assessment may lack contextual accuracy, reinforcing the need for a complementary AI–teacher assessment model. These findings suggest a pedagogical implication shift toward training EFL students to use AIGSA not merely as a simple answer tool but rather as a metacognitive scaffold. EFL instructors are advised to adopt a complementary model in which students are trained to integrate AIGSA for low-stakes practice while relying on instructor feedback for high-stakes assessment. Adopting such a strategy fosters students' metacognitive skill development and develops their judgment for navigating AI tools autonomously and critically.

Conclusion

The study examined EFL students' perceptions in two underrepresented contexts regarding the use of AIGSA in their learning and its statistical relationship with metacognitive skill development. The findings revealed that EFL undergraduates demonstrated a high level of engagement in using AIGSA in their learning. This indicates frequent acceptance of AI-based self-assessment as a learning support tool, integrated by students into their independent and lifelong learning practices. The results identified a statistically significant moderate positive correlation between using AIGSA and self-reported enhancement in metacognitive skills among students. This suggests that using AIGSA stimulates students' engagement in monitoring, reflecting, and evaluating their learning strategies.

Additionally, the study showed that nationality was associated with significant differences in students' reported perceptions toward AIGSA, with Saudi EFL students reporting higher perceptions (with a small effect size) than their Yemeni counterparts. This difference may be attributed to variations in digital infrastructure and training that incorporate AI tools in their learning.

Furthermore, the findings also highlight EFL undergraduates' metacognitive agency in strategically selecting feedback sources based on task demands and learning goals. The findings showed perceived affordances of AIGSA, rather than as direct indicators of assessment validity. Building on Al Harrasi's (2024) emphasis on teacher-supported evaluative knowledge, the present findings extend self-assessment and self-directed learning literature by demonstrating that AIGSA can provide complementary scaffolding that supports autonomous monitoring and evaluation. The study contributes to the literature of AI and self-assessment in EFL learning by introducing the term "Artificial Intelligence Generated Self-Assessment" (AIGSA); a term where students use accessible AI tools to assess their language learning. Additionally, the current study's strongest contribution is that learners view AIGSA as useful for low-pressure practice, clarification, and self-monitoring, while still trusting instructors more for holistic and performance-based assessment. The importance of raising students' awareness of the limits of self-assessment, underscoring a complementary AI–teacher assessment model, in which trust, support, and validity are negotiated across feedback sources. The current findings extend existing literature by demonstrating that AIGSA contributes meaningfully to EFL undergraduates' metacognitive skill development while its perceived validity remains situational and complementary to teacher assessment. A major contribution of this study lies in its mixed-approach, which combines quantitative evidence of a positive relationship between AIGSA and metacognitive skills with qualitative insights into how learners view trust, support, and validity across AI-generated and teacher feedback.

The findings set important theoretical and practical implications. Theoretically, the study extends existing literature on self-assessment and metacognition by conceptualizing AIGSA as a learner-centered strategy which helps learners engaging in monitoring, evaluating, and reflecting on their learning experience. Pedagogically, the findings suggest that EFL instructors can effectively integrate AIGSA into formative assessment and independent learning practice to promote students' metacognitive awareness and self-regulated capacities. However, the results indicate that EFL students continue to value their teacher assessment for communicative skills and high-stakes evaluation. Consequently, this finding supports the complementary assessment framework where AIGSA is employed in low-stakes practices while teacher assessment is utilized in holistic-sensitive evaluation.

For learners, AIGSA supports autonomous and lifelong learning by encouraging continuous reflection beyond the classroom and strategy adjustment. At the institutional level, course designers, policymakers, and AI developers should consider incorporating AI literacy and AI-based assessment into language programs and aligning them into curricular goals and providing supportive learning environments.

Given that self-report measures to collect quantitative and qualitative findings may not fully capture actual strategic and metacognitive behavior, future research should employ longitudinal and experimental designs that triangulate learner perceptions with behavioral and observational data. Further studies are needed to examine how sustained engagement with AIGSA supports the development of evaluative knowledge and self-efficacy, and how these capacities facilitate a gradual shift toward autonomous self-assessment within self-directed learning frameworks across diverse EFL contexts, including the role of teacher assessment literacy in AI-supported practices.

Funding

This work was funded by the Deanship of Scientific Research, Vice Presidency for Graduate Studies and Scientific Research, King Faisal University, Saudi Arabia [Grant number- KFU263964].

Acknowledgements

The authors extend sincere gratitude to the Deanship of Scientific Research, Vice Presidency for Graduate Studies and Scientific Research, King Faisal University, for funding this project.

References

- Al Harrasi, K. T. S. (2024). Enhancing learner self-monitoring in self-assessment through the use of pedagogical resources. *Cogent Social Sciences*, 10(1), 2334486. <https://doi.org/10.1080/23311886.2024.2334486>
- Al-Hoorie, A. H., & AlShakhori, M. (2025). ChatGPT in language learning and teaching. In A. H. Al-Hoorie, C. Mitchell, & T. Elyas (Eds.), *Language education in Saudi Arabia: Integrating technology in the classroom* (pp. 147–166). Springer Nature Switzerland.
- Alkhateeb, A., Hezam, A. M. M., & Almuraikhi, A. A. (2025). Assessing the use of AI tools for EFL exam preparation at Saudi universities: Efficiency, benefits, and challenges. *Cogent Education*, 12(1), 2507553. <https://doi.org/10.1080/2331186X.2025.2507553>
- Alshaie, F. S., Bin-Hady, W. R. A., Rashed, R. Q. G., & Madani, O. (2025). The impact of EFL instructors' AI competence on developing the quality of English education in Ha'il: Teachers' perspective. *Cogent Education*, 12(1). <https://doi.org/10.1080/2331186X.2025.2585559>
- Altındağ, M., & Senemoğlu, N. (2013). Metacognitive skills scale. *Hacettepe Üniversitesi Eğitim Fakültesi Dergisi*, 28(28–1), 15–26.
- Andrade, H. L. (2019). A critical review of research on student self-assessment. *Frontiers in Education*, 4, 1–13. <https://doi.org/10.3389/feduc.2019.00087>

- Anggoro, K. J., & Pratiwi, D. I. (2023). Fostering self-assessment in English learning with a generative AI platform: A case of quizizz AI. *Studies in Self-Access Learning Journal*, 14(4). <https://doi.org/10.37237/140406>
- Berg, J. A., Repstad, Ø., Serkland, T. T., Mork, T. E., Bru, C., Tenstad, O., & Kvernenes, M. (2025). Student perspectives on AI-supported formative assessment in pharmacology. *British Journal of Clinical Pharmacology*, 1–8. <https://doi.org/10.1002/bcp.70267>
- Bernard, R. C., & Kermarrec, G. (2022). Peer assessment and video feedback for fostering self, co, and shared regulation of learning in a higher education language classroom. *Frontiers in Education*, 7, Article 732094. <https://doi.org/10.3389/feduc.2022.732094>
- Bhatti, J. A., Ilyas, F., Naseer, H., & Farooq, M. (2025). Exploring the impact of AI- powered assessment tool on test anxiety and academic self-efficacy. *Review of Applied Management and Social Sciences*, 8(2), 1007–1021. <https://doi.org/10.47067/ramss.v8i2.533>
- Bin-Hady, W. R. A., Ali, J. K. M., & Al-humari, M. A. (2024). The effect of ChatGPT on EFL students' social and emotional learning. *Journal of Research in Innovative Teaching & Learning*, 17(2), 243–255. <https://doi.org/10.1108/JRIT-02-2024-0036>
- Bin-Hady, W. R. A., Al-Kadi, A., Hazaea, A., & Ali, J. K. M. (2025). Exploring the dimensions of ChatGPT in English language learning: A global perspective. *Library Hi Tech*, 43 (4–5), 1315–1328. <https://doi.org/10.1108/LHT-05-2023-0200>
- Bin-Hady, W. R. A., Al-kadi, A., Alzubi, A. A., & Mahdi, H. S. (2020). Assessment of language learning strategies in technology-enhanced learning environment. In S. Meri Yilan & K. Koruyan (Eds.), *ICT-based assessment, methods, and programs in tertiary education* (pp. 83–99). IGI Global Scientific Publishing. <https://doi.org/10.4018/978-1-7998-3062-7.ch005>
- Braun, V., & Clarke, V. (2006). Using thematic analysis in psychology. *Qualitative Research in Psychology*, 3(2), 77–101. <https://doi.org/10.1191/1478088706qp0630a>
- Braun, V., & Clarke, V. (2019). Reflecting on reflexive thematic analysis. *Qualitative Research in Sport, Exercise and Health*, 11(4), 589–597. <https://doi.org/10.1080/2159676X.2019.1628806>
- Braun, V., & Clarke, V. (2020). Can I use TA? Should I use TA? Should I not use TA? Comparing reflexive thematic analysis and other pattern-based qualitative analytic approaches. *Counselling and Psychotherapy Research*, 21(1), 37–47. <https://doi.org/10.1002/capr.12360>
- Chen, S.-Y. R. (2026). Two voices, one draft: A diachronic comparison of human and AI feedback in EFL essay writing. *Australian Journal of Applied Linguistics*, 9, 103200. <https://doi.org/10.29140/ajal.2026.103200>
- Chen, A., Zhang, Y., Jia, J., Liang, M., Cha, Y., & Lim, C. P. (2025). A systematic review and meta-analysis of AI-enabled assessment in language learning: Design, implementation, and effectiveness. *Journal of Computer Assisted Learning*, 41(1), e13064. <https://doi.org/10.1111/jcal.13064>
- Conway-Smith, B., West, R. L., & Mylopoulos, M. (2023). Metacognitive skill: How it is acquired. In *Proceedings of the Annual Meeting of the Cognitive Science Society* (Vol. 45, No. 45).
- Creswell, J. W., & Clark, V. L. P. (2017). *Designing and conducting mixed methods research*. Sage Publications.
- Damayanti, A. M., Syamsiyah, N., Astuti, E., Dania, U., & Kusumaningtyas, P. (2021). The role of meta-cognitive skills in developing the 21st century skills. *Educational Studies: Conference Series*, 1(1), 26–30. <https://doi.org/10.30872/escs.v1i1.856>
- Daniar, A. V., Herdyastuti, N., & Lutfi, A. (2023). Analysis effectiveness of implementation assessment as learning on metacognitive skills. *IJORER: International Journal of Recent Educational Research*, 4(6), 759–770.
- Ertel, W. (2024). *Introduction to artificial intelligence*. Springer Nature.
- Farooq, U., Malik, S. K., & Syed, A. S. (2024). Revolutionizing higher education: The transformative impact of artificial intelligence on teaching, learning, and career. *ShodhKosh: Journal of Visual and Performing Arts*, 5(3), 789–794. <https://doi.org/10.29121/shodhkosh.v5.i3.2024.3239>
- Flavell, J. H. (1979). Metacognition and cognitive monitoring: A new area of cognitive developmental inquiry. *American Psychologist*, 34(10), 906–911.
- Gortaire-Díaz, D., Almache-Granda, G., Real-Roby, R., & Mora-Herrera, E. (2026). Bridging the gap: How AI tools support self-directed EFL skill development inside the classroom. *Revista InveCom*, 6(2), 1–14. <https://zenodo.org/records/15788013>

- Henderson, M., Bearman, M., Chung, J., Fawns, T., Buckingham Shum, S., Matthews, K. E. & de Mello Heredia, J. (2025). Comparing generative AI and teacher feedback: Student perceptions of usefulness and trustworthiness. *Assessment & Evaluation in Higher Education*, 1–16. <https://doi.org/10.1080/02602938.2025.2502582>
- Hooda, M., Rana, C., Dahiya, O., Rizwan, A., & Hossain, M. S. (2022). Artificial intelligence for assessment and feedback to enhance student success in higher education. *Mathematical Problems in Engineering*, 2022(1), 5215722. <https://doi.org/10.1155/2022/5215722>
- Khlaif, Z. N., Ayyoub, A., Hamamra, B., Bensalem, E., Mitwally, M. A., Ayyoub, A., Hattab, M., & Shadid, F. (2024). University teachers' views on the adoption and integration of generative AI tools for student assessment in higher education. *Education Sciences*, 14(10), 1090. <https://doi.org/10.3390/educsci14101090>
- Klyshbekova, M., & Abbott, P. (2024). ChatGPT and assessment in higher education: A magic wand or a disruptor? *Electronic Journal of e-Learning*, 22(2), 30–45. <https://doi.org/10.34190/ejel.21.5.3114>
- Lavrysh, Y., Leshchenko, M., & Tymchuk, L. (2023). Development of metacognitive skills through digital narratives in higher education. *Revista Electrónica de Investigación Educativa*, 25. <https://doi.org/10.24320/redie.2023.25.e07.5028>
- Lee, A. V. Y. (2023). Supporting students' generation of feedback in large-scale online course with artificial intelligence-enabled evaluation. *Studies in Educational Evaluation*, 77, 101250. <https://doi.org/10.1016/j.stueduc.2023.101250>
- Metwally, A. A., & Bin-Hady, W. R. A. (2025). Probing the necessity and advantages of AI integration training for EFL educators in Saudi Arabia. *Cogent Education*, 12(1), 2472462. <https://doi.org/10.1080/2331186X.2025.2472462>
- Marantika, J. E. R. (2021). Metacognitive ability and autonomous learning strategy in improving learning outcomes. *Journal of Education and Learning (EduLearn)*, 15(1), 88–96. <http://doi.org/10.11591/edulearn.v15i1.17392>
- Munaroh, N. L. (2024). Asesmen dalam Pendidikan: Memahami Konsep, Fungsi dan Penerapannya. *Dewantara: Jurnal Pendidikan Sosial Humaniora*, 3(3), 281–297. <https://doi.org/10.30640/dewantara.v3i3.2915>
- Nachar, N. (2008). The Mann-Whitney U: A test for assessing whether two independent samples come from the same distribution. *Tutorials in Quantitative Methods for Psychology*, 4(1), 13–20. <https://doi.org/10.20982/tqmp.04.1.p013>
- Najaf, A. R. E., Agussalim, & Sihananto, A. N. (2025). Optimizing AI ChatGPT: Prompt engineering in the learning assessment process for high school students. *Nusantara Science and Technology Proceedings*, 2024(47), 543–550. <https://doi.org/10.11594/nstp.2025.4780>
- Nazaretsky, T., Mejia-Domenzain, P., Swamy, V., Frej, J., & Käser, T. (2026). Who gives feedback matters: Student biases towards human and AI-Generated formative feedback. *Journal of Computer Assisted Learning*, 42(1), e70153.
- Nieminen, J. H., Yan, Z., & Boud, D. (2025). Self-assessment design in a digital world: Centring student agency. *Assessment & Evaluation in Higher Education*, 50(5), 732–746. <https://doi.org/10.1080/02602938.2025.2467647>
- Njoku, C., & Wey-Amaewhule, B. (2025). Integration of artificial intelligence (AI) driven assessment tools on the academic performance of students in public universities in Rivers State. *International Journal of Educational Management*, 1(1), 1–16. <https://ijedm.com/index.php/ijedm/article/view/2>
- Ratnayake, A., Bansal, A., Wong, N., Saseetharan, T., Prompiengchai, S., Jenne, A., Thiagavel, J., & Ashok, A. (2024). All “wrapped” up in reflection: Supporting metacognitive awareness to promote students' self-regulated learning. *Journal of Microbiology and Biology Education*, 25(1), e00103–23. <https://doi.org/10.1128/jmbe.00103-23>
- Register, A. D. (2026). Humanizing the machine student psychological responses to AI-generated feedback in online education. In J. Smith (Ed.), *Psychology, ethics, and AI in online higher education* (pp. 289–318). IGI Global Scientific Publishing.
- Sahmaniasl, R. (2025). EFL educators' perceptions of AI-driven assessments: A systematic review. *International Journal of Multidisciplinary Research and Growth Evaluation*, 6(3), 672–680.
- Seema, P.V., & Padmanabha, C.H. (2024). Conceptual framework on metacognitive skills. *i-manager's Journal on Educational Psychology*, 17(4), 54–64. <https://doi.org/10.26634/jpsy.17.4.20649>

- Slimi, Z., Benayoune, A., & Alemu, A. E. (2025). Students' perceptions of artificial intelligence integration in higher education. *European Journal of Educational Research*, 14(2), 471–484. <https://doi.org/10.12973/eu-jer.14.2.471>
- Sultana, S. S., Renugadevi, R., Bhargavi, M., & Biyabani, S. A. A. (2025). AI-driven evaluation techniques: Revolutionizing student practices. In J. Smith (Ed.), *Adopting artificial intelligence tools in higher education* (pp. 1–22). CRC Press.
- Teng, M. F., & Ma, M. (2024). Assessing metacognition-based student feedback literacy for academic writing. *Assessing Writing*, 59, 100811. <https://doi.org/10.1016/j.asw.2024.100869>
- Veenman, M. V. (2017). Assessing metacognitive deficiencies and effectively instructing metacognitive skills. *Teachers College Record*, 119(13), 1–20. <https://doi.org/10.1177/016146811711901303>
- Veenman, M. V. J. (2011). Alternative assessment of strategy use with self-report instruments: A discussion. *Metacognition and Learning*, 6(2), 205–211. <https://doi.org/10.1007/s11409-011-9080-x>
- Voss, E. (2024). Language assessment and artificial intelligence. In A. J. Kunnan (Ed.), *The concise companion to language assessment* (pp. 112–125). John Wiley & Sons.
- Yan, Z., Panadero, E., Wang, X., & Zhan, Y. (2023). A systematic review on students' perceptions of self-assessment: Usefulness and factors influencing implementation. *Educational Psychology Review*, 35(3), 81. <https://doi.org/10.1007/s10648-023-09799-1>
- Zheng, C., Huang, Z., Ma, S., & Ma, X. (2024, October). SelfGauge: An intelligent tool to support student self-assessment in GenAI-enhanced project-based learning. In *Adjunct Proceedings of the 37th Annual ACM Symposium on User Interface Software and Technology* (pp. 1–3).
- Zimmerman, B. J. (2002). Becoming a self-regulated learner: An overview. *Theory into Practice*, 41(2), 64–70. https://doi.org/10.1207/s15430421tip4102_2

Appendix A

Spearman Correlation Coefficients for Item-Dimension Associations

Items	Correlation (*r*)	Significance (*p*)
1. Interactive and engaging AI assessments increase my motivation to learn English.	.714**	.000
2. AI-based self-assessment enhances my engagement and participation in English language learning activities.	.680**	.000
3. Using AI-generated assessment encourages me to take control of my own learning process.	.669**	.000
4. Using AI-generated feedback help me adjust my English learning strategies and improve performance.	.744**	.000
5. Using AI-generated assessment enhances my achievement and grades.	.764**	.000
6. AI helps me track my progress and set realistic goals for learning English.	.734**	.000
7. AI-powered assessments have reduced my anxiety and stress when evaluating my English language skills.	.440**	.000
8. I use my previous learning experiences while organizing my new learning.	.531**	.000
9. If the learning couldn't be accomplished, I search for other strategies that could be effective.	.625**	.000
10. I revise my study plan that I used in learning and make necessary corrections.	.675**	.000
11. When a learning strategy that I used fails in learning process, I employ new one.	.688**	.000
12. I search for the reasons of the failure while learning a subject.	.735**	.000
13. I search for how I learned a subject most effectively while learning.	.768**	.000
14. I revise and correct the learning strategies while studying a subject.	.767**	.000
15. I assess if the cognitive strategy that I employ has been successful or not.	.615**	.000
16. I plan how and when to use the resources that will help me learn a subject well.	.417**	.000
17. I don't have an exact idea of how to organize my learning.	.522**	.000
18. While learning a subject, I am not aware of employing which strategy and how to use it.	.429**	.000
19. I know how much time I need to learn a subject.	.467**	.000

(Continued)

Items	Correlation (*r*)	Significance (*p*)
20. I have difficulty in planning my learning a subject in accordance with my own learning qualities.	.527**	.000
21. I have difficulty in distinguishing important parts about a text or a learning unit.	.537**	.000
22. I prepare the learning environment that is necessary for learning process.	.574**	.000
23. I critically make a plan before beginning to study a text.	.415**	.000
24. I know which subjects I can learn easily and which I will have difficulty in learning.	.534**	.000
25. I know the other subject matters that I can use an effective learning strategy in a subject.	.541**	.000
26. I determine which learning strategy I should e.	.425**	.000
27. I know when I need to ask for help.	.501**	.000
28. I determine what I will learn about a subject before I start studying it.	.426**	.000
29. It is important for me to overview my learnings from time to time to determine how much and what I learned.	.455**	.000
30. I recognize my errors during learning process.	.441**	.000
31. I check if I understood a subject during learning.	.553**	.000
32. I have difficulty in understanding the reason of the trouble I experienced during learning.	.525**	.000
33. I check if I effectively use my time during learning.	.479**	.000
34. It is important for me to build meaningful relations between learned subjects during learning.	.375**	.000
35. Till I reach a result, I organize the conditions for keeping my attention.	.391**	.000
36. I don't spare much time for monitoring how much I learned about the subject during learning process.	.245**	.000
37. During learning process, I have difficulty to determine in which conditions I can learn and those I have failed to learn.	.496**	.000