EFL University Lecturers' Perceptions of AI and Critical Thinking: Opportunities, Boundaries, and Assessment Dilemmas

Khuloud Alouzi*[1], Dafa Allah Ibrahim[2], Sara Omran[3], Salem Aladi[4], Ahmed Hamid Ahmed[5]

[1,2,3,4]University of Zawia, Libya
[5]Alzaiem Alazhari University, Sudan
*Correspondence: ✉***k.alozzi@zu.edu.ly**

Abstract

This study explored EFL university lecturers' perceptions of AI tools in relation to critical thinking, focusing on perceived opportunities, acceptable-use boundaries, and assessment dilemmas at the University of Zawia, Libya. The study addresses a gap in research because evidences from Libyan universities remain limited, especially on how lecturers set boundaries and protect assessment validity when AI is widely available. A mixed-methods descriptive approach was used, combining a questionnaire (N = 70) and semi-structured interviews (n = 10) to link survey patterns with lecturers' practical explanations and examples. Survey results showed moderate endorsement of AI opportunities for supporting critical thinking (M = 3.93, SD = 1.08), especially for language and comprehension support (M = 4.23, SD = 0.97) and higher-order question generation (M = 4.03, SD = 1.05). However, lecturers strongly emphasized boundaries for acceptable AI use (M = 4.42, SD = 0.89), particularly the need for an institutional policy (M = 4.60, SD = 0.73) and disclosure of AI use (M = 4.49, SD = 0.86). Assessment dilemmas were high (M = 4.17, SD = 0.98), with strongest concern about plagiarism/patchwriting risks (M = 4.40, SD = 0.84) and difficulty judging students' own critical thinking (M = 4.24, SD = 0.97). Interview findings reinforced a "benefit-boundary tension," highlighting preference for assessment redesign over detection-based policing. As a specific contribution, the study proposes an "Opportunity - Boundary - Assessment redesign" model that translates lecturers' perceptions into practical actions for EFL departments. The study recommends regulated use guidance, staff training, and student integrity support.

INTRODUCTION

Critical thinking is widely recognized as a key outcome of higher education because it supports reasoned judgment, evidence-based decision-making, and responsible participation in academic and social life (Abdulghani et al., 2025; Ibrahim et al., 2025). In EFL contexts, critical thinking is also connected to core language-learning goals such as interpreting texts, evaluating arguments, and producing well-justified academic writing (Ennis, 2011; Hilman et al., 2024; Syarifaturrahmatullah et al., 2025). However, developing critical thinking through English is not always straightforward, because learners must manage language demands (vocabulary, grammar, and discourse) at the same time as they manage complex reasoning and interpretation. For this reason, EFL lecturers often face a dual responsibility where they are supporting language development while also promoting higher-order thinking such as analysis, evaluation, and argumentation (Bloom et al., 1956). In practice, this means that thinking quality in EFL is often judged by language products (essays, reports, responses, etc.), which makes assessment vulnerable when language can be generated or heavily assisted by AI tools.

In recent years, the rapid spread of generative artificial intelligence (AI) tools has added a new layer of complexity to EFL teaching and assessment. Generative AI systems can produce fluent explanations, summaries, questions, and essay-style responses, which may support learning when used as scaffolding, but may also create risks when outputs are treated as "ready-made" answers (Kasneci et al., 2023; Setiawan et al., 2023; Kasheem, M. et al., 2025). In EFL classrooms, this issue is especially sensitive because language accuracy and academic style can be generated quickly, which can hide the real level of students' understanding and reasoning. As a result, lecturers may experience a tension between using AI as learning support and protecting the validity of assessment and the authenticity of student work (Messick, 1995). Recent research in educational technology argues that, instead of focusing only on detection, institutions should pay attention to pedagogy, transparency, and assessment redesign to maintain academic integrity and meaningful learning outcomes (Bearman et al., 2024; Alouzi, 2024; Al Dokali et al., 2025). This concern is also reflected in international scholarship which discusses how generative AI affects assessment design, evaluative judgment, and academic integrity (Bearman et al., 2024; Zawacki-Richter et al., 2019), and in empirical work warning that AI-generated fluency can distort teachers' judgments of learning evidence in higher education contexts (Kasneci et al., 2023).

From a learning perspective, AI tools may reduce linguistic and cognitive barriers by assisting with vocabulary explanations, paraphrasing, and background knowledge, which can free students' attention for reasoning and interpretation (Sweller, 1988; Vygotsky & Cole, 1978). For example, if learners receive support in understanding a difficult reading passage, they may become more able to ask questions, consider alternative interpretations, and justify claims. At the same time, critical thinking requires more than producing fluent language; it requires careful reasoning, selecting evidence, and reflecting on uncertainty (Facione, 1990). When students rely heavily on AI outputs, there is a risk of superficial engagement, weak evidence use, or patchwriting practices that threaten academic integrity and reduce the visibility of students' own thinking processes (Bearman et al., 2024; Baroud, 2024; Almajri et al., 2025). Therefore, the key question is not whether AI can support language learning, but whether EFL assessment can still capture students' reasoning and evidence use in a trustworthy way.

These issues are particularly relevant in the Libyan university context, where EFL lecturers work with students who often have different levels of English proficiency and limited prior experience with academic argumentation in English. In such conditions, tools that support language and comprehension may appear attractive, but assessment becomes more challenging because lecturers must judge whether assignments represent students' own critical thinking or AI-assisted production. Assessment theory emphasizes that the meaning of scores depends on the quality of evidence collected and the extent to which tasks truly represent the intended construct (Messick, 1995). Therefore, lecturers' perceptions are important not only as attitudes, but also as indicators of how teaching and assessment may change in response to AI. Moreover, institutional policy and training availability vary across contexts, so the same AI tool may produce different outcomes depending on local support, enforcement, and assessment culture.

Although research on AI in education is expanding, there is still a practical need to understand how EFL lecturers, as key decision-makers, perceive AI in relation to critical thinking, boundaries of acceptable use, and assessment dilemmas, especially in under-researched contexts. Teacher perceptions matter because they influence classroom rules, task design, feedback practices, and the implementation of academic integrity expectations (Zawacki-Richter et al., 2019; Yahya et al., 2025; Wiresti et al., 2025). However, current discussions often remain general and do not sufficiently describe how

lecturers negotiate AI use in daily teaching and assessment, particularly within specific institutional and cultural environments. Moreover, many studies report "AI benefits and risks" broadly, but provide limited detail on (a) how lecturers translate perceived benefits into concrete boundary rules, and (b) what assessment redesign strategies are preferred when integrity and validity become uncertain.

Accordingly, the novelty of the present study is that it offers a context-specific, lecturer driven account of AI use in EFL that integrates three elements in one model: perceived opportunities, preferred boundaries, and assessment redesign responses. In the Libyan university context, the rapid and often unregulated spread of generative AI creates a practical problem for EFL teaching and assessment: while AI can support comprehension and classroom engagement, it also makes it harder for lecturers to determine whether submitted work represents students' own critical thinking or AI-assisted production. This situation may threaten assessment validity, fairness, and academic integrity, particularly in writing-based tasks where fluent language can mask limited reasoning (Boud & Falchikov, 2006; Messick, 1995; Paisun et al., 2024).

At the University of Zawia, there is still limited empirical evidence on how EFL lecturers understand these opportunities and risks and how they set boundaries in real classroom practice. Therefore, this study addresses the following research questions: (1) What opportunities do EFL lecturers perceive for generative AI to support students' critical thinking in EFL learning? (2) What boundaries do lecturers consider acceptable versus unacceptable for AI use in EFL courses? (3) What assessment dilemmas do lecturers report regarding authorship, validity, and academic integrity in AI-supported student work? (4) How do interview explanations help interpret and contextualize the survey patterns within the University of Zawia setting?

For this reason, the present study investigates EFL university lecturers' perceptions at the University of Zawia, Libya, focusing on perceived opportunities, acceptable-use boundaries, and assessment dilemmas. Using a mixed-methods descriptive approach (questionnaire and interviews), the study aims to provide practical evidence to inform department-level policy, teacher training priorities, and student guidance for responsible AI use in EFL learning and assessment (Masoud et al., 2025; Maati et al., 2025; Abrahem & Baroud, 2025). As a specific contribution to the current scholarship and scholarly debates, the study proposes an integrated "Opportunity-Boundary-Assessment

redesign" framing that can guide future research and help EFL institutions move from general debates about AI to actionable, assessment-focused policy and classroom practice.

RESEARCH METHODOLOGY

This study employed a mixed-methods descriptive design with an explanatory sequential (quantity to quality) to examine EFL university lecturers' perceptions of AI and critical thinking at the University of Zawia, Libya. The target population was EFL lecturers teaching English-related courses (e.g., writing, reading, literature, linguistics) across relevant faculties at the university. A non-probability sampling approach was used because the study aimed to capture accessible lecturers while ensuring variation in teaching roles. Specifically, convenience sampling was used for questionnaire distribution, supported by purposive inclusion to ensure representation of different course types and experience levels.

The final survey sample included 70 lecturers, selected based on the criteria that they (a) teach EFL courses at the University of Zawia and (b) have current involvement in student assessment. Data were first collected through a questionnaire that contained two sections: (a) background items (teaching experience, main course taught, AI-use frequency, policy awareness, and training experience) and (b) 12 closed-ended Likert items measured on a 5-point scale (1 = strongly disagree to 5 = strongly agree). The Likert items were organized into three domains (4 items each): Opportunities, Boundaries, and Assessment dilemmas, and the item wording focused on classroom and assessment situations to make responses comparable across lecturers. The questionnaire was administered online (Google Forms) and remained open for approximately 2–3 weeks, and responses were screened for completeness before analysis. Questionnaire data were analyzed using descriptive statistics (frequencies and percentages for categorical variables; means and standard deviations for each item and domain score). Domain scores were computed by averaging the four item means within each domain, to provide a clear summary of lecturers' overall stance in each area.

Moreover, to enrich and explain the survey patterns, the second phase used semi-structured interviews with 10 lecturers from the same university. Interview participants were selected using purposive (maximum-variation) sampling from those who agreed to be interviewed, to reflect differences in years of experience, course type, and AI engagement (e.g., frequent vs limited

AI use). The interviews explored lecturers' understandings of critical thinking in EFL, concrete classroom examples of AI benefits and risks, boundary-setting practices (what they allow and prohibit), and assessment challenges (e.g., take-home writing credibility, redesigning tasks, and disclosure expectations). Each interview lasted approximately 30–45 minutes, was conducted in English, and was audio-recorded with informed consent before transcription. Interview data were analyzed using thematic analysis with a hybrid coding approach: initial codes were developed from the study domains (opportunities, boundaries, dilemmas), while additional codes were allowed to emerge inductively from the data. To support replicability and trustworthiness, the study used (a) an interview protocol with the same core questions for all participants, (b) a clear coding procedure documented by code definitions, and (c) representative quotes linked to each theme in the Results section.

Validity and reliability procedures were applied to strengthen the quality of findings. For the questionnaire, content validity was supported through expert review (two EFL lecturers and one educational measurement specialist) who checked item clarity, domain alignment, and relevance to EFL assessment. A small pilot test (n ≈ 10 lecturers) was conducted to confirm that items were understandable and to refine wording before final distribution. Internal consistency reliability was checked using Cronbach's alpha for each domain, and items were retained because they showed acceptable consistency for a short scale (α ≥ .70 as a minimum guideline). For the qualitative phase, credibility was supported through member-checking of brief summaries (participants confirmed the accuracy of main points) and peer debriefing between the researchers during theme refinement. Finally, integration occurred at the interpretation stage by comparing survey patterns with interview themes to explain why lecturers support AI opportunities while also emphasizing strong boundaries and assessment safeguards.

RESULT AND DISCUSSION
RESULT
Participants' profile and AI-use background

This subsection describes the lecturer sample and their reported AI-use context at the University of Zawia (Libya). The questionnaire was completed by 70 EFL university lecturers, representing a range of teaching experience and course responsibilities. Descriptive statistics (frequencies and percentages) were used to summarize background variables in a transparent way. Overall, the

profile indicates that AI is already present in lecturers' professional environment, but formal policy awareness and structured training remain limited, which can explain later concerns about fairness and assessment validity. This background is important since perceptions of AI often depend on institutional support and prior exposure; therefore, interpretation of later findings should consider that many lecturers may be making judgments under policy uncertainty rather than under clear official guidance (Zawacki-Richter et al., 2019; Bearman et al., 2024; Masoud et al., 2025).

Table 1. Participant profile and AI-use background (N = 70)

Variable	Category	n	%
Years of teaching experience	1–5 years	14	20.0
	6–10 years	19	27.1
	11–15 years	17	24.3
	16+ years	20	28.6
Main course taught (primary)	Writing / Academic writing	21	30.0
	Reading / Comprehension	18	25.7
	Literature	16	22.9
	Linguistics / Applied linguistics	15	21.4
AI-use frequency (any purpose)	Never	9	12.9
	Rarely (monthly or less)	16	22.9
	Sometimes (weekly)	28	40.0
	Often (several times/week)	17	24.3
Awareness of university/department AI policy	Aware of a clear policy	18	25.7
	Not aware / unclear	52	74.3
Received training on AI for teaching/assessment	Yes	15	21.4
	No	55	78.6

As shown in Table 1, most lecturers reported moderate to long teaching experience, suggesting that perceptions reflect established teaching practice rather than only early-career views. Course responsibilities were diverse (writing, reading, literature, linguistics), which is important because AI risks and acceptable-use boundaries often differ by course type, especially in writing-based assessment. Regarding AI engagement, a majority of lecturers indicated at least occasional AI use for teaching-related tasks (e.g., preparing materials or generating prompts), while a smaller group reported frequent use. However, awareness of an institutional or departmental AI policy was not high, and only a minority reported receiving formal training. This distribution supports the

interpretation that lecturers' strong desire for boundaries and policy (reported in later subsections) is partly shaped by the absence of consistent guidance and shared enforcement mechanisms. At the same time, an alternative explanation is that some lecturers may underreport policy awareness or training because policies are informal or communicated inconsistently across units; therefore, these results should be read as "perceived clarity" rather than confirmed institutional absence.

Opportunities: How lecturers think AI can support critical thinking in EFL
***Quantitative results***

Lecturers at the University of Zawia reported a moderately positive view of AI as a tool that can support critical thinking in EFL learning. The overall Opportunities domain score was M = 3.93 (SD = 1.08), indicating that most respondents tended to agree that AI has pedagogical value, although the level of agreement was not uniform across items. In particular, the highest endorsement was for OPP2 (AI supports language so students can focus on reasoning) with M = 4.23 (SD = 0.97). This suggests that lecturers strongly recognize AI as a linguistic scaffold in EFL contexts where language difficulty can limit students' ability to engage in deeper interpretation and evaluation. The second strongest item was OPP1 (AI helps generate higher-order questions) with M = 4.03 (SD = 1.05), indicating perceived usefulness for initiating inquiry and guiding discussion around texts and ideas.

Table 2 Opportunities (AI & Critical Thinking)

Code	Statement	Mean	SD
OPP1	AI helps students generate higher-order questions about texts (e.g., "why/how" questions).	4.03	1.05
OPP2	AI supports language (vocabulary/grammar) so students can focus more on reasoning and interpretation.	4.23	0.97
OPP3	AI helps students consider alternative interpretations and counterarguments.	3.84	1.14
OPP4	AI prompts/feedback encourage students to justify claims using textual evidence.	3.60	1.17

Figure 1 Opportunities (AI & Critical Thinking)
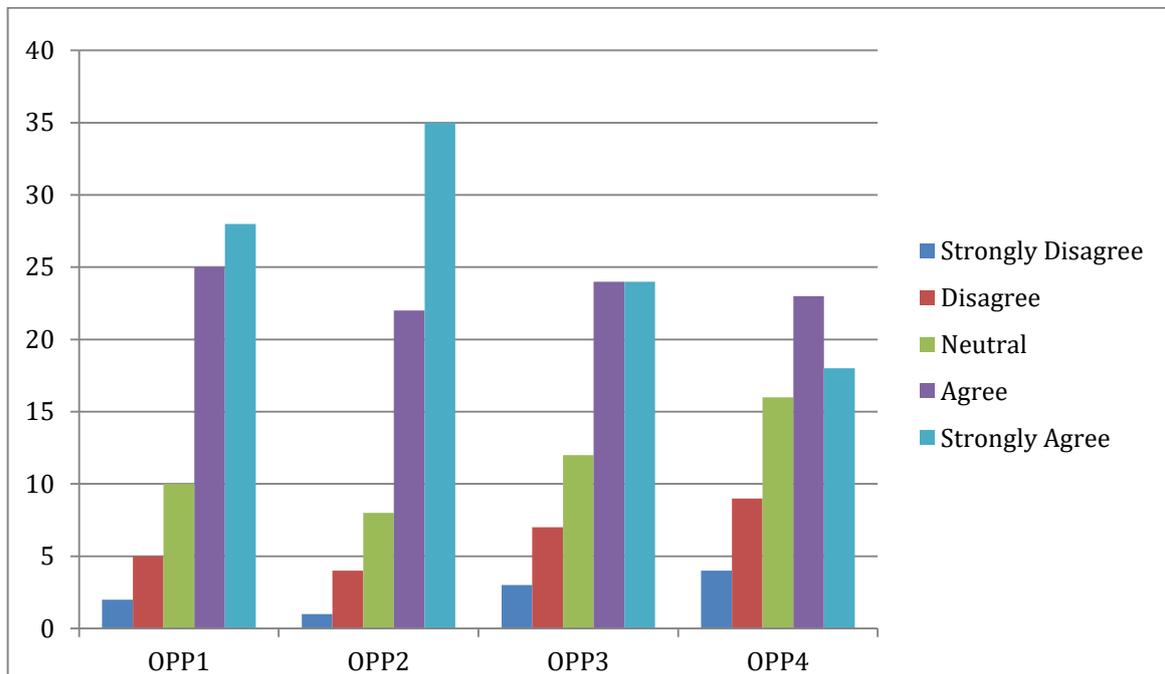
However, lecturers were more cautious regarding AI's role in strengthening text-based justification. The lowest opportunity item was OPP4 (AI encourages students to justify claims using textual evidence) with M = 3.60 (SD = 1.17). This pattern shows an important distinction in lecturers' thinking. AI is viewed as helpful for supporting access and generating ideas, but less trusted for building the discipline of evidence-based reasoning that is central to critical thinking in academic literacy (Primarni et al., 2025; Shalghoum et al., 2025; Muttaqin et al., 2026). In other words, lecturers' survey responses suggest that AI may help students "start thinking," but does not automatically ensure that students will verify, cite, and justify their interpretations responsibly.

Qualitative interview themes

Interview findings clarified why lecturers rated certain opportunities more strongly than others. Under Theme A (AI as scaffolding), lecturers frequently described AI as useful for vocabulary explanations, background context, and question prompts that help students enter the text and reduce comprehension barriers. Under Theme B (AI as idea expander), lecturers noted that AI can provide alternative interpretations and counterarguments, which can enrich classroom discussions, but only when the teacher sets clear limits and asks students to evaluate and refine outputs. At the same time, Theme C (reduced productive struggle) appeared across interviews. Lecturers worried

that students may accept AI responses as "final answers," which can weaken the intellectual struggle that is often necessary for critical reading and meaning-making in literature and academic texts (Masuwd et al., 2024; Pallawagau et al., 2025; Manshur et al., 2025). This concern integrates with the broader debates that AI may encourage surface-level completion unless teachers explicitly require evaluation, comparison, and justification steps (Bearman et al., 2024). Still, this theme reflects lecturers' professional judgment rather than direct observation of students' cognitive processes, so it should be interpreted as perceived risk rather than proven learning decline.

DISCUSSION

The dominance of OPP2 (language support) is expected in EFL settings because linguistic difficulty often increases cognitive load and reduces learners' capacity to engage in higher-level reasoning. When AI assists with vocabulary, grammar, and background explanations, students can allocate more attention to analysis, questioning, and evaluation—processes closely associated with critical thinking (Sweller, 1988; Masuwd, 2024; Ayad et al., 2025). In this sense, lecturers' perceptions align with the view of tools as scaffolds that extend learners' current performance, especially when tasks are above the learner's independent language ability (Vygotsky & Cole, 1978).

Similarly, higher endorsement of OPP1 suggests that lecturers value AI for initiating inquiry through question generation, which can support dialogic learning and classroom discussion when questions are used as prompts rather than as substitutes for interpretation. This interpretation is consistent with previous literature that describes AI as potentially useful for low-stakes learning support and ideation, but not automatically reliable for deep reasoning or academic judgement (Kasneci et al., 2023; Zawacki-Richter et al., 2019). At the same time, a limitation is that our survey did not measure students' proficiency or course level, which may moderate how useful AI scaffolding is in different classes.

At the same time, the relatively lower endorsement of OPP4 indicates lecturers' concern that AI does not guarantee disciplined evidence use. Evidence-based reasoning requires selecting relevant textual proof, connecting it logically to claims, and acknowledging uncertainty—skills that may be weakened if students rely on AI-generated explanations without verifying sources or quoting the text accurately (Facione, 1990; Abdullaha et al., 2026). Lecturers in interviews repeatedly implied that AI can "sound academic" even

when reasoning is not grounded in the text, which creates a gap between fluent language and genuine critical engagement. Therefore, a practical implication is that lecturers tend to position AI as helpful mainly in pre-thinking stages, such as pre-reading support, planning, question generation, and brainstorming, while they are more cautious about AI involvement in final interpretation and argument justification. This suggests the need for regulated instructional routines (e.g., "AI for preparation, but human evidence for claims"), so that AI support does not replace the core practices of critical thinking in EFL literacy.

Boundaries: What lecturers consider acceptable vs. unacceptable AI use
***Quantitative results (boundary strength and policy priorities)***

Lecturers expressed strong agreement that AI use in EFL should be governed by clear boundaries. The overall Boundaries domain score was M = 4.42 (SD = 0.89), indicating a high level of consensus that regulation is necessary. The most strongly endorsed item was BOUND4 (need for an institutional policy) with M = 4.60 (SD = 0.73). This suggests that lecturers do not view AI governance as an individual teacher responsibility only; rather, they expect departmental or university-level guidance to ensure consistency across courses and assessments. The second highest item was BOUND2 (students should disclose AI use) with M = 4.49 (SD = 0.86), showing that transparency is perceived as central for maintaining academic trust. These findings reflect the concerns that policy clarity is necessary to reduce inconsistency and to support fair assessment decisions when generative AI becomes normal in student work. Still, strong agreement here may also reflect uncertainty and risk perception rather than direct experience with misconduct cases.

Table 3 Boundaries (acceptable use)

Code	Statement	Mean	SD
BOUND1	Students may use AI for brainstorming, but not for producing final answers/interpretations.	4.39	0.92
BOUND2	Students should disclose AI use when submitting assignments.	4.49	0.86
BOUND3	AI use should be limited during graded assessments (e.g., exams, take-home tests).	4.20	1.04
BOUND4	Teachers need a clear institutional policy on acceptable AI use.	4.60	0.73

Figure 3 Boundaries (acceptable use)
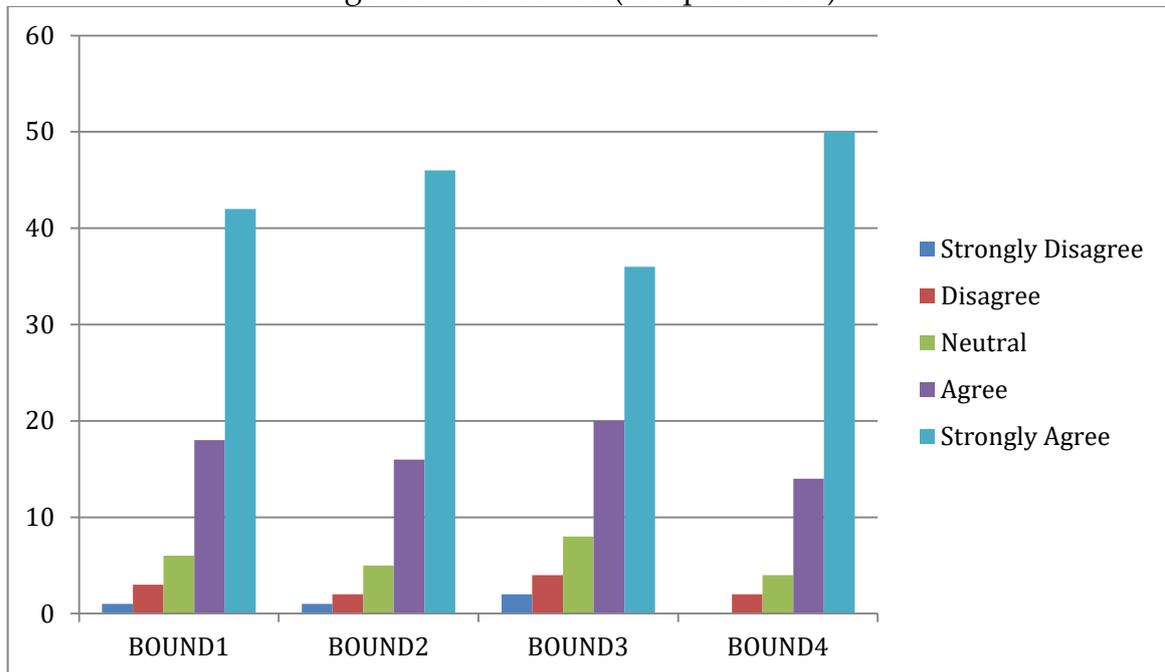
A clear "regulated-use" orientation was also visible in BOUND1, where lecturers supported allowing AI for early-stage support (e.g., brainstorming) but not for final interpretations, with M = 4.39 (SD = 0.92). This pattern indicates that lecturers differentiate between process support (permitted) and product substitution (not permitted). In practical terms, lecturers appear to endorse AI as a tool that can assist learning activities, while insisting that assessed outcomes must still represent the student's own reasoning and expression.

Qualitative interview themes (how boundaries are understood)

Interview responses explained the strong boundary scores by highlighting professional concerns around authorship and assessment credibility. Under Theme D (support, not authorship), lecturers repeatedly described AI as acceptable when it supports comprehension, planning, or practice, but unacceptable when it replaces students' work or voice. Under Theme E (need for shared rules and training), lecturers emphasized that informal, teacher-by-teacher rules are not enough; they requested a shared departmental policy and student training to reduce confusion and to promote fair enforcement. Finally, Theme F (course-type differences) showed that lecturers' strictness is not uniform across the curriculum: boundaries were described as stricter in writing and assessment-heavy courses, compared with

literature or reading discussion courses where AI might be used more flexibly for preparation and support.

Discussion (integrity, uncertainty, and regulated-use policies)

The strong endorsement of boundaries can be interpreted as a response to uncertainty about authorship and the increased risk of academic misconduct in AI-mediated learning. When lecturers cannot clearly distinguish students' own reasoning from AI-generated text, assessment validity is threatened, particularly in writing-based outcomes where language quality can mask limited understanding (Boud & Falchikov, 2006). In this study, lecturers' emphasis on institutional policy (BOUND4) reflects a desire to protect fairness and standardization, which are key principles in responsible assessment practice (Brown, & Abeywickrama, 2019; Alrumayh, 2025). Put simply, lecturers seem to be saying, without shared rules, both teaching and grading become inconsistent. This interpretation is in accordance with scholarly literature which emphasizing that generative AI challenges traditional assumptions about authorship and requires assessment redesign rather than reliance on detection only (Bearman et al., 2024).

Disclosure (BOUND2) emerged as a preferred mechanism for transparency and trust, and this aligns with broader integrity approaches that emphasize openness and ethical authorship practices. However, lecturers also recognized a practical tension: disclosure is difficult to verify if students do not report honestly, and teachers often lack reliable tools to prove AI use. This challenge supports the growing view that assessment redesign and clear task constraints may be more effective than detection-based policing (Bearman et al., 2024). Overall, the survey and interview evidence suggest that lecturers lean toward a regulated-use policy, where AI is permitted for specific functions (e.g., brainstorming, language support, question prompts) but prohibited for outcomes that substitute student authorship (e.g., final interpretations, complete essays). Such a policy approach keeps AI within a learning-support role while attempting to protect the integrity of critical thinking assessment in EFL contexts. However, we should acknowledge a limitation here. The strict boundaries may reduce equitable access to learning support if some students depend on AI for language learning; therefore, policy should balance integrity protection with inclusive learning support, especially for lower-proficiency learners.

Assessment dilemmas: Why AI creates grading and validity concerns
***Quantitative results (level and pattern of concern)***

Lecturers reported high concern about assessment in the presence of AI, as reflected in the Assessment dilemmas domain score (M = 4.17, SD = 0.98). The most strongly endorsed dilemma item was DILEM2 (AI increases the risk of plagiarism/patchwriting) with M = 4.40 (SD = 0.84), indicating that many lecturers view AI as intensifying textual borrowing problems and reducing confidence in written submissions. The second highest item was DILEM1 (AI makes it difficult to judge students' own critical thinking) with M = 4.24 (SD = 0.97).

This suggests that the main worry is not only "cheating" as a moral issue, but also a measurement problem—lecturers are uncertain whether written work still provides valid evidence of students' reasoning and interpretation skills. This result is consistent with assessment literature that treats validity as evidence-based inference; when evidence is contaminated by external assistance, score interpretation becomes less defensible (Messick, 1995). However, our survey cannot determine whether these concerns differ by discipline, course level, or class size, which may shape lecturers' assessment confidence.

Table 4 Assessment dilemmas

Code	Statement	Mean	SD
DILEM1	AI makes it difficult to judge students' own critical thinking in assignments.	4.24	0.97
DILEM2	AI increases the risk of plagiarism/patchwriting in writing tasks.	4.40	0.84
DILEM3	Current assessment methods are less valid when AI is widely used by students.	3.89	1.07
DILEM4	I need training to assess critical thinking fairly in AI-supported student work.	4.16	1.02

Figure 4 Assessment dilemmas
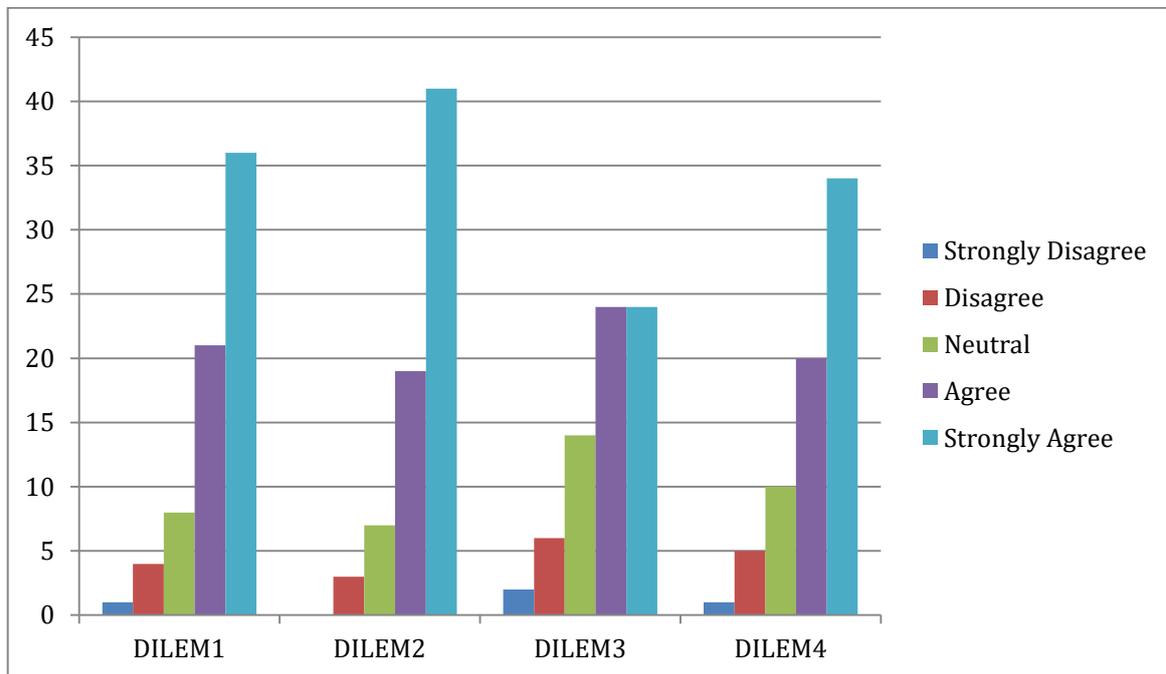
Although slightly lower, the concern about overall assessment validity remained strong. DILEM3 (current assessment methods are less valid when AI is widely used) had M = 3.89 (SD = 1.07), showing that lecturers tend to agree that traditional take-home assignments may no longer represent student ability as before. Finally, the high mean for DILEM4 (need training) (M = 4.16, SD = 1.02) indicates that lecturers do not only identify the problem; they also acknowledge a professional need for practical support to redesign assessment and to evaluate critical thinking fairly in AI-supported work. An alternative explanation is that "training need" may reflect rapid technological change rather than lack of competence. Lecturers may be requesting institutional integration, shared rubrics, and examples of AI-resilient tasks more than general AI awareness training.

Qualitative interview themes (how dilemmas are experienced)

Interviews provided strong explanations for the survey pattern. Under Theme G (unclear student thinking), lecturers repeatedly expressed uncertainty about authorship and reasoning ownership, especially in assignments that look "too fluent" compared to students' known level. Under Theme H (limits of detection), lecturers emphasized that AI-detection tools and plagiarism systems are not fully reliable, which creates anxiety and sometimes conflict with students. As a result, many participants shifted the focus from policing to

redesign. Under Theme I (assessment redesign suggestions), lecturers proposed practical alternatives such as in-class writing, oral defense, and process portfolios (drafts, reflections, and evidence trails) to make student thinking visible and assessable.

Discussion (validity threats and assessment shift)

The results point to a central assessment problem: AI disrupts authorship, and therefore changes what student work can validly represent. In assessment theory, validity depends on whether evidence collected from tasks supports appropriate interpretations of learner ability (Messick, 1995; Hadi & Masuwd, 2025). When students can outsource large parts of drafting and reasoning to AI, a traditional take-home essay may no longer measure the intended construct (critical thinking in EFL), but instead measures access to tools and ability to prompt or edit outputs. This explains why DILEM1 and DILEM3 remain high: lecturers are concerned that assessment results may not reflect students' real reasoning, interpretation, and evidence use. Still, we should avoid claiming that take-home essays are "invalid" in all cases. The results suggest that lecturers perceive reduced confidence in validity when task conditions and authorship are unclear.

The high endorsement of training needs (DILEM4) suggests that lecturers do not feel fully equipped to respond with reliable assessment practices. This is consistent with recent arguments that higher education should focus less on detection and more on integrity-driven assessment design, including tasks that elicit authentic performance and verifiable processes (Bearman et al., 2024; Kasheem, A, et al., 2025). Based on both the survey and interviews, a key implication is a shift toward process evidence (drafting histories, justification notes, reflective commentary, and citation checks) and performance evidence (in-class writing, oral defense, short interpretive tasks completed under supervision). Such approaches can maintain academic standards while still allowing limited AI use for learning support, especially at preparatory stages. In this way, assessment can remain aligned with critical thinking outcomes while recognizing the new reality of AI access in EFL contexts.

Integrating survey and interviews: the "benefit–boundary tension" model

When survey and interview findings are considered together, a consistent pattern emerges that can be described as a benefit–boundary tension. Quantitatively, lecturers reported moderate agreement that AI can support critical thinking in EFL (Opportunities: M = 3.93, SD = 1.08), while they also

endorsed very strong boundaries on acceptable AI use (Boundaries: M = 4.42, SD = 0.89) and expressed high assessment concerns (Dilemmas: M = 4.17, SD = 0.98). This combined profile suggests that lecturers do not reject AI as a learning tool; instead, they accept its usefulness mainly for preparation and scaffolding, while simultaneously expressing a strong need for control to protect assessment credibility.

The interviews explained this tension clearly. Lecturers described AI as helpful for reducing language barriers and generating questions (learning benefits), but also as a source of uncertainty about authorship and evidence of learning (assessment threat). Such a pattern is consistent with assessment scholarship emphasizing that technology adoption in education must be matched with assessment designs that maintain valid inferences about students' abilities (Boud & Falchikov, 2006; Husin et al., 2025). Thus, this model should not be interpreted as evidence that AI necessarily harms critical thinking; it shows that lecturers perceive both benefits and risks, and they prioritize assessment credibility when stakes are high.

A practical way to represent this integration is a short "translation model" that links perceived opportunities to concrete safeguards. For example, when lecturers see AI as useful for vocabulary and context, they prefer a boundary that limits AI to pre-reading support, and they recommend assessment redesigns such as in-class interpretation tasks or brief oral explanation to confirm understanding.

Similarly, when AI is used for idea generation and counterarguments, lecturers tend to require disclosure and to prohibit full AI-written submissions, while suggesting process-based portfolios (drafts + reflection) to make student reasoning visible. Overall, the central interpretive claim of this study is that AI is acceptable when it supports learning steps, but lecturers want clear safeguards to protect (a) evidence of critical thinking and (b) academic integrity, especially in writing-based assessments. This aligns with recent arguments that institutions should prioritize integrity-driven assessment redesign over detection-only approaches (Bearman et al., 2024; Elihami et al., 2024; Sulistyowati et al., 2025).

Practical implications for the University of Zawia

The findings imply the need for a structured and shared response at the department level, rather than individual teacher solutions. First, the English department (or faculty) can introduce a simple policy template that includes: (1) a disclosure statement (students indicate when and how AI was used), (2) a list

of allowed uses (e.g., vocabulary support, background summaries, question prompts, outlining), and (3) prohibited outputs (e.g., AI-generated full essays, final interpretations submitted as student work, fabricated references, or copied AI text without citation). Such policy clarity can reduce inconsistency across courses and improve fairness, which lecturers emphasized as a major concern in interviews and which is important for sustainable assessment practice (Brown & Abeywickrama, 2019; Alrumayh et al., 2025; Masuwd, 2026; Hajijah et al., 2025). Policy should also include brief examples of acceptable vs unacceptable use, because students often interpret rules differently when guidance is general. At the same time, policy should avoid overly broad prohibitions that may push AI use underground; a realistic policy may work better when it combines permission (for learning support) with accountability (disclosure and evidence-based requirements).

Second, lecturers' high agreement on training needs suggests three immediate professional development priorities: (a) prompt literacy for pedagogical use (designing prompts that require reasoning and evidence rather than producing final answers), (b) verification routines (teaching and assessing textual evidence use, source checking, and identifying hallucinated claims), and (c) AI-resilient assessment design, such as oral defenses, timed in-class writing, process portfolios, and reflective commentaries that document reasoning steps. These approaches are consistent with assessment perspectives that emphasize aligning tasks with learning outcomes and capturing evidence of thinking, not only final products (Boud & Falchikov, 2006; Messick, 1995; Hasibuan et al., 2024). Third, students need clear guidance to develop responsible AI practices in EFL: short training sessions or course handouts can focus on academic integrity, citation and acknowledgement of assistance, fact-checking, and reflective use (e.g., students explain what they accepted/rejected from AI outputs and why). In this way, the University of Zawia can benefit from AI as a learning support tool while protecting critical thinking outcomes and maintaining assessment credibility.

CONCLUSION

This study examined how EFL lecturers at the University of Zawia (Libya) perceive generative AI in relation to critical thinking and assessment. Overall, lecturers recognized meaningful opportunities for AI to support EFL learning, mainly by reducing language barriers and helping students generate questions and alternative viewpoints. This pattern is consistent with

international discussions that generative AI can function as learning scaffolding and idea-generation support, particularly in language-mediated tasks. At the same time, lecturers strongly supported clear boundaries, especially disclosure of AI use and the development of institutional policy, indicating that AI acceptance is conditional rather than open-ended. Such boundary emphasis aligns with assessment scholarship that highlights the need for clear rules and shared standards to protect fairness and credible score interpretation when authorship becomes uncertain. Importantly, assessment concerns were high: lecturers expressed that AI makes it difficult to judge students' real critical thinking and increases the risk of plagiarism and patchwriting, which can weaken the validity of traditional take-home assignments. Therefore, the main conclusion is that AI is not simply a "tool to adopt" in EFL; it changes what counts as evidence of learning, and it pushes departments to rethink assessment conditions and documentation of student reasoning.

The integrated findings suggest that the key issue is not only whether AI can be useful, but how EFL departments can protect evidence of learning while allowing limited, pedagogically meaningful use. Beyond the University of Zawia, these findings have broader implications for similar higher education contexts where policy guidance and staff training may be developing: universities may need regulated-use frameworks that define allowed AI functions (support) while safeguarding prohibited outcomes (authorship substitution). In response, the study supports a regulated-use approach in which AI is permitted for preparation (e.g., vocabulary support, context, outlining, and question prompts) but restricted for final assessed products unless clearly acknowledged and accompanied by process evidence. This conclusion moves the discussion from general debates about "AI benefits/risks" toward a practical assessment-focused direction, where validity is strengthened through process evidence (drafts, justification notes, reflection) and performance checks (in-class writing and oral defense). Practical steps for the University of Zawia include establishing departmental guidelines, providing teacher training in AI-resilient assessment (e.g., in-class tasks, oral defense, portfolios, and reflective justifications), and offering students explicit instruction on integrity, citation, and verification. These measures can help maintain critical thinking outcomes in EFL while responding realistically to the growing presence of generative AI in university learning. Future research can extend this work by testing whether these assessment redesign strategies

improve the quality of evidence for critical thinking and reduce integrity concerns across different EFL course types and student proficiency levels.

REFERENCES

Abdulghani, N., Masuwd, M., Alrumayh, S., Masoud, M., & Touré, Y. (2025). Maqasid al-Shariah as a Framework for Developing Critical Thinking in Islamic Higher Education. *Journal of Islamic Studies and Social Science*, 1(2), 47-63.

Abdullaha, L., Baroud, N., Alsaeh, F., kasheem, A., Alouzi, K., Shalghoum, N., & Abdullah, M. (2026). Advancing Sustainable Development Goals through Islamic Education: A Mixed-Methods Study among Scientific Disciplines at the University of Zawia. *Amorti: Jurnal Studi Islam Interdisipliner*, 5(1), 42–58. doi:10.59944/amorti.v5i1%20Januari.589

Abrahem, M. ., & Baroud, N. . (2025). EXPLORING CULTURAL, POLITICAL, AND SOCIOECONOMIC INFLUENCES ON COLLECTIVE PSYCHOLOGICAL RESILIENCE IN LIBYA. *Jurnal Ilmu Psikologi Dan Kesehatan (SIKONTAN)*, 3(4), 163–176. doi:10.47353/sikontan.v3i4.2784

Al Dokali, N., Aljarmi, A., & Baroud, N. (2025). AI Applications in Chemistry Education: Student Engagement, Learning Outcomes, and Practical Insights. *LAVOISIER: Chemistry Education Journal*, 4(2), 140-154. doi:10.24952/lavoisier.v4i2.17475

Almajri, S., Baroud, N., Alouzi, K. M., & Kasheem, A. (2025). Islamic psychology: An integrative approach to human behavior and mental well-being. *Bulletin of Islamic Research*, 3(4), 687–704. doi:10.69526/bir.v3i4.353

Alouzi, K. M. (2024). Assessment of Requirements for the Creation of Course Materials for English as a Second Language for Students Majoring in Subjects Other Than English. *Majapahit Journal of English Studies*, 2(1), 19–34. doi:10.69965/mjes.v2i1.105

Alrumayh, S. (2025). AI and Qur'anic Interpretation: Exploring the Ethical and Epistemological Boundaries of Artificial Intelligence in Understanding the Qur'an. *Al Furqan: Jurnal Ilmu Al Quran Dan Tafsir*, 8(2), 223-239. doi:10.58518/alfurqan.v8i2.4243

Alrumayh, S., Ayad, N., Alouzi, K., Ibrahim, D., Abdullah, M., Masoud, M., & Kasheem, M. (2025). Perceptions of Islamic Studies, Sharia, and Law Students Towards the Use of Artificial Intelligence in English Learning. *Action Research Journal Indonesia (ARJI)*, 7(3), 2238 – 2256. doi:10.61227/arji.v7i3.504

Ayad, N., Masuwd, M. A., & Alrumayh, S. (2025). From riba to zakat: An analytical study of Islamic economic principles and their distinction from conventional economics. *Bulletin of Islamic Research*, 3(4), 733–752. doi:10.69526/bir.v3i4.358

Baroud, N. (2024). A Systematic Comparison of Students Attitudes Toward Practical Work in Chemistry Department Faculty of Education, Zawia - University of Zawia: A Systematic Comparison of Students Attitudes Toward Practical Work . *IJCER (International Journal of Chemistry Education Research)*, 8(2), 137–143. doi:10.20885/ijcer.vol8.iss2.art7

Bearman, M., Tai, J., Dawson, P., Boud, D., & Ajjawi, R. (2024). Developing evaluative judgement for a time of generative artificial intelligence. *Assessment & Evaluation in Higher Education*, 49(6), 893-905. doi:10.1080/02602938.2024.2335321

Bloom, B. S., Engelhart, M. D., Furst, E. J., Hill, W. H., & Krathwohl, D. R. (1956). *Taxonomy of educational objectives: The classification of educational goals. Handbook 1: Cognitive domain* (pp. 1103-1133). New York: Longman.

Boud, D., & Falchikov, N. (2006). Aligning assessment with long-term learning. *Assessment & evaluation in higher education*, 31(4), 399-413. doi:10.1080/02602930600679050

Brown, H. D., & Abeywickrama, P. (2019). Language Assessment: Principles and Class-Room Practices (3rd Ed.). Pearson Education.

Creswell, J. W., & Clark, V. L. P. (2017). *Designing and conducting mixed methods research*. Sage publications.

Elihami, E., Masuwd, M. A., Sheerly, S., Ismail, I., Sitonda, S., & Sudirman, M. Y. (2024). Data-driven approaches in Islamic quality management and education technology for advancing sustainable development goals. *Jurnal Pendidikan Progresif*, 14(3), 1599–1616. doi:10.23960/jpp.v14.i3.2024109

Ennis, R. H. (2011). The nature of critical thinking: An outline of critical thinking dispositions and abilities. *University of Illinois*, 2(4), 1-8.

Facione, P. (1990). Critical thinking: A statement of expert consensus for purposes of educational assessment and instruction (The Delphi Report).

Hadi, N., & Masuwd, M. A. (2025). Classical Cooperative Learning Model for Reading Classic Literature: Enhancing Student Independence Through Self-Regulation. *Izdihar: Journal of Arabic Language Teaching, Linguistics, and Literature*, 8(1). doi:10.22219/jiz.v8i1.36829

Hajijah, A., Nasution, M., Hasibuan, N. A. P., Hilda, L., & Baroud, N. Y. (2025). Fostering critical thinking through socio-scientific issue-based problem-based learning in stoichiometry instruction. *JTK (Jurnal Tadris Kimiya)*, 10(1), 35–44. doi:10.15575/jtk.v10i1.43992

Hasibuan, A., Nasution, I. F. A., & Masuwd, M. (2024). Tarekat in the digital age: Transforming spirituality for the age of technology. *Religia*, 27(1), 13–30. doi:10.28918/religia.v27i1.2306

Hilman, C., Ishom, M., Sarkawi, S., & Alouzi, K. M. (2024). Adoption of mobile-assisted language learning in improving college students' English listening skills. *International Journal of Teaching and Learning*, 2(10), 2711–2722.

Husin, H., Aziz, A. bin A., & Masuwd, M. (2025). Integrating Al-Ghazali's educational philosophy: Advancing transformative learning in Islamic schools in the digital era. *SYAMIL: Journal of Islamic Education*, 13(1), 29–51. doi:10.21093/sy.v13.i1.10263

Ibrahim, D., Aboujanah, Y., Alouzi, K., Albshkar, H., Masoud, M., & Almajri, S. (2025). The use of artificial intelligence-based translation tools for language department students. *Journal of Arabic Literature, Teaching and Learning*, 1(3), 76–92. Retrieved from https://jaliter.intischolar.id/index.php/jaliter/article/view/7

Kasheem, A., Baroud, N., & Almajri, S. K. (2025). The role of emotional intelligence in enhancing teacher-student relationships in rural schools. *SIBATIK Journal: JurnalIlmiahBidangSosial, Ekonomi, Budaya, Teknologi, Dan Pendidikan*, 4(8), 1693–1704. doi:10.54443/sibatik.v4i8.3158

Kasheem, M., Yahya, N., Shalghoum, N., Masuwd, M., Alriteemi, A., Abdullah, M., Alsaeh, F., & Alrumayh, S. (2025). Artificial intelligence in academic research: Adoption, opportunities, and barriers among faculty in Libya higher education. *Multidisciplinary Journal of Thought and Research*, 1(3), 109–127. Retrieved from https://mujoter.intischolar.id/index.php/mujoter/article/view/20

Kasneci, E., Seßler, K., Küchemann, S., Bannert, M., Dementieva, D., Fischer, F., … & Kasneci, G. (2023). ChatGPT for good? On opportunities and challenges of large language models for education. *Learning and individual differences*, 103, 102274.

Maati, A., Alzletni, N., Yahya, N., Barkah, S., Aladi, S., Alrumayh, S., … Masuwd, M. (2025). Bridging Faith and Sustainability: Faculty Attitudes toward Integrating Research Priorities with Islamic Educational Values and SDG 4. *Journal of Multidisciplinary Research of Education*, 1(3), 178–190. doi:10.34125/jomre.v1i3.34

Manshur, U., Hadi, N., & Masuwd, M. A. (2025). Classical cooperative learning model for reading classic literature: Enhancing student independence through self-regulation. *Izdihar: Journal of Arabic Language Teaching, Linguistics, and Literature*, 8(1), 105–120. doi:10.22219/jiz.v8i1.36829

Masoud, M., Kasheem, M., Barkah, S., Alsaeh, F., Baroud, N., & Albshkar, H. (2025). Balancing Technology and Empathy: Faculty Perceptions of Artificial Intelligence in University Counseling. *Coution: Journal Counseling and Education*, 6(2), 147-166. doi:10.47453/coution.v6i2.3624

Masuwd, M. (2026). Revitalizing the Mosque's Role as an Economic Stabilizer for the Muslim Community: Social Fiqh Perspectives and Contemporary Practice. *Al Qalam: Jurnal Ilmiah Keagamaan dan Kemasyarakatan*, 20(1), 291–307. doi:10.35931/aq.v20i1.5913

Masuwd, M. A. (2024). Islamic Pluralism and Küng's Global Ethical Discourse: Toward a "Global-Maqāṣid" Centered Paradigm. *Ittesaal–Journal of Connecting Discourses*, 41-63. doi:10.64984/ijcd.1.2.2024.03

Masuwd, M., Sumanik, E. D., Sarkawi, S., & Amer, M. A. B. (2024). Measuring foreign language anxiety: Concerning students' motivation and their self-perception. *Int. J. Teach. Learn*, 2(8), 2087-2099.

Messick, S. (1995). Validity of psychological assessment: Validation of inferences from persons' responses and performances as scientific inquiry into score meaning. *American psychologist*, 50(9), 741.

Muttaqin, M. Z., Alrumayh, S. H., & Barkah, S. A. M. (2026). Utilization of Social Media in Living Qur'an Learning: An Analysis of Generation Z Students. *Tunjuk Ajar: Journal of Education and Culture*, 2(1), 1–27. doi:10.64929/ta.v1i2.34

Paisun Paisun, Syarifah, & Masuwd, M. (2024). Investigating The Relationship Between Teacher Self-Efficacy and Student Achievement. *Andragogi: Jurnal Pendidikan Dan Pembelajaran*, 4(2), 94–108. doi:10.31538/adrg.v4i2.1303

Pallawagau, B., Masuwd, M., Alsayd, A. D., Albshkar, H. A., & Alzletni, N. S. (2025). Nasruddin Umar's love curriculum and its relation to Arabic language teaching: Socio-linguistic and semantic studies / منهجالحبعندنصرالدينعمروعلاقتهبتعليماللغةالعربية: دراساتاجتماعيةلغويةودلالية. *Ijaz Arabi Journal of Arabic Learning*, 8(2). doi:10.18860/ijazarabi.v8i2.32264

Paul, R., & Elder, L. (2019). *The miniature guide to critical thinking concepts and tools*. Rowman & Littlefield.

Primarni, A., Masuwd, M., Makmudi, M., Fa'atin, S., & Nuhdi, A. (2025). The implementation of the theoretical humanism paradigm in holistic education. *Journal of Research in Instructional*, 5(1), 199-209. doi:10.30862/jri.v5i1.601

Setiawan, D., Masuwd, M. A., Maliki, N., Laily, I. F., & Fitriyani, Y. (2023). Impact of Digital Storytelling for Developing Oral Communication Skills, Digital Literacy, and Learning Motivation Among Pre-service Elementary Teachers. *International Journal of Educational Qualitative Quantitative Research*, 2(2), 34–42. doi:10.58418/ijeqqr.v2i2.118

Shalghoum, N., Yahya, N., Abdullah, M., Masuwd, M., Kasheem, M., Alrumayh, S., Aryanti, Y., & Rosyadi, N. (2025). Integrating Maqasid al-Shariah into Higher Education: Enhancing the Role of Faculty in Achieving the SDGs. *International Journal of Islamic Studies Higher Education*, 4(2), 153–173. doi:10.24036/insight.v4i2.231

Sulistyowati, E., Masnun, M. A., & Yahya, N. (2025). An Ideal Model for The Preparation of Internal Regulations of Universities in Indonesia Based on Meaningful Participation in The Success of Sustainable Development Goals (Sdg's) 2030. *Jurnal Suara Hukum*, 7(2). doi:10.26740/jsh.v7n2.p463-487

Sweller, J. (1988). Cognitive load during problem solving: Effects on learning. *Cognitive science*, 12(2), 257-285.

Syarifaturrahmatullah, S., Ahmad Fadhel Syakir Hidayat, Adamu Abubakar Muhammad, Afrah Masood, & Mowafg Masuwd. (2025). Exploration and Mapping of Methods for Researching Second Language Classrooms: Content Analysis based on Rod Ellis' book. *El-Syaker : Samarinda International Journal of Language Studies*, 2(2), 73–95. doi:10.64093/esijls.v2i2.524

Vygotsky, L. S., & Cole, M. (1978). *Mind in society: Development of higher psychological processes*. Harvard university press.

Wiresti, R. D., Wibowo, D. V., Kuswanto , A. V., & Sami Barkah. (2025). Islamic parenting education: Emotional stimulation of children in the family. *At Turots: JurnalPendidikan Islam*, 7(2), 1416–1426. doi:10.51468/jpi.v7i2.1224

Yahya, N., Abdullah, M., & Masuwd, M. (2025). Development of Digital Education in Libya: Progress, Challenges, and Future Directions. *International Journal of Education and Digital Learning (IJEDL)*, 3(5), 211–219. doi:10.47353/ijedl.v3i5.307

Zawacki-Richter, O., Marín, V. I., Bond, M., & Gouverneur, F. (2019). Systematic review of research on artificial intelligence applications in higher education–where are the educators?. *International journal of educational technology in higher education*, 16(1), 1-27.