Vol. 2, No. (8) Sep. 2026

ISSN: 4556-3489

Arab Journal of Advanced Research and Development (AJARD)

المجلة العربية للبحوث المتقدمة والتنمية

The Influence of Generative AI on EFL Assessment Practices: Lecturers' Perceptions, Challenges, and Responses at the University of Benghazi

Mohammed Moussa Mohamme

University of Benghazi, Faculty of Arts, Al Marj Campus, Al Marj, *Correspondence*

Abstract

English language assessment in higher education now faces new challenges due to the rapid development and widespread use of generative artificial intelligence (AI). This mixed-methods study looks at how English as a foreign language instructor have reacted to the increasing use of generative AI and how they perceive its impact on student assessments. A questionnaire was used to gather information from thirty English language instructors, and eight of the respondents participated in semi-structured interviews. The results indicate that lecturers' opinions of the validity and reliability of present assessment methods, especially take-home assignments and unsupervised written work, have been greatly impacted by generative AI. Writing skills were thought to be especially susceptible to the use of generative AI, and participants voiced worries about differentiating students' authentic work from AI-assisted writing. The results also indicate that instructors are starting to reevaluate their methods of assessment, focusing more on oral, process-based, explanatory, and in-class assessments that let students directly demonstrate their knowledge and comprehension. However, the interviews uncovered real-world obstacles to assessment reform, such as big class sizes, time constraints, and a lack of explicit institutional guidelines and rules. Another issue that emerged was equitable access to AI tools. Overall, the study indicates that generative AI is transforming EFL assessment procedures and that instructors require clear policies, training, and institutional support to maintain academic integrity while modifying assessments to fit the evolving educational landscape.

Keywords: Academic Integrity; College-level English instruction; EFL Assessment; Generative Artificial Intelligence; Instructors Assessment Practices.

1. Introduction

For a long time, language assessment relied on traditional methods that measured individual learners' abilities in writing and expressing their comprehension of the English language; however, the emergence of generative AI has allowed any student to create a complete essay or assignment without exerting effort, leading to a lack of reliability in assessments. AI technologies have led to a significant advancement in the field of language teaching, particularly for EFL learners, as they can either assist or compete with students in completing assignments and homework; consequently, teachers are compelled to develop new types of questions and create different methods of assessment and grading to uphold academic integrity and thoroughly evaluate assignments. Traditional assessment methods are no longer enough to ensure the measurement of the student's real level.

Despite the massive growth in the research that discusses the general challenges of AI, there's a noticeable scarcity of studies examining the practical and field responses of the teachers

0096893649662 | info@ajard.org | www.ajard.org

themselves; most previous studies focused on the students' perspectives or plagiarism detection, while there's still a clear gap about how EFL teachers adapt their testing practices as a response to these technologies. Accordingly, this research specifically seeks to investigate EFL teachers' perspectives on generative AI tools and monitor actual modifications and changes that they have introduced into their methods to design assignments and help students keep up with this technological development. The significance of this study is that it helps teachers in academic departments improve their assessment methods and preserve academic integrity. This research concludes that the widespread diffusion of AI requires a complete transformation in teachers' practices and necessitates changes in their traditional methods, offering a broader perspective on how teachers should respond to this technology.

1.2. Research Question

This research addresses three core questions:

1. What are EFL teachers' perceptions of the impact of generative AI on English language assessment?

2. To what extent has generative AI influenced EFL teachers' testing and assessment practices?

3. What changes have EFL teachers made to their assessment methods in response to generative AI?

2. Literature Review

2.1. Introduction and background about generative AI

Traditional assessment methods are no longer enough to ensure the student's real level. Despite the massive growth in research that discusses the general challenges of AI, there's a noticeable scarcity of studies examining the practical and field responses of the teachers themselves; most previous studies focused on the students' perspectives or plagiarism detection, while there's still a clear gap in how EFL teachers adapt their testing practices as a response to these technologies. Accordingly, this research specifically seeks to investigate EFL teachers' perspectives on generative AI tools and monitor actual modifications and changes that they have introduced into their methods to design assignments and help students keep up with this technological development. The significance of this study is that it helps teachers in academic departments improve their assessment methods and preserve academic integrity. This research concludes that the widespread diffusion of AI requires a complete transformation in teachers' practices and necessitates changes in their traditional methods, offering a broader perspective on how teachers should respond to this technology.

2.2. Theoretical frameworks and teachers' perceptions

It is necessary to understand that teachers' perspective toward generative artificial intelligence exerts a profound impact on their propensity to utilize it and manage it. The Technology Acceptance Model (TAM) can assist in identifying those perceptions based on technological usefulness in education (Albion & Ertmer, 2002; Becker *et al*., 1999). Theoretically, teachers' beliefs and their usual behavior form the basis for acceptance of artificial intelligence tools (Hogg

& Vaughan, 2005). As provided by Wozney & Abramy (2006), a model offers an interpretation for the variance in teachers' acceptance of those technologies and clarifies the divergence of opinions between advocates of the idea, citing ease of use and integration in the syllabus, and opponents to avoid complete dependence upon them. As artificial intelligence enables students to generate complete texts and solve any type of exams and home tasks, teachers face a major obstacle due to this problem, represented by the difficulty of fair evaluation (Gardner *et al*., 1993).

This compels teachers to design assessment tasks differently and newly, requiring conscious human thinking to combat the complete dependence on those technologies and preserve evaluation integrity. The technical competence of teachers must be taken into account to establish a connection between the aforementioned points. The reluctance of less technologically competent teachers to integrate technology is a result of the most common obstacles to rejecting the use of technology in language education, as stated by Al-Oteawi (2002) and Zammit (1992). Additionally, teachers who are more technically knowledgeable exhibit increased self-assurance and a more effective ability to incorporate it into the curriculum (Saaid, 2010).

2.3. Teacher Perceptions of Generative AI and Its Impact on Assessment Practice

Some members of the educational staff view these technologies as beneficial supplements for English language teaching, considering them digital assistants that provide immediate feedback to students (Chiu, 2023). In contrast, the opposing side warns that these technologies will generate a decline and reduction in students' abilities that require practice, which will threaten independent student writing (Cotton *et al*., 2023). The capacity of artificial intelligence to generate fluent texts compels teachers to change their evaluation metrics (Eaton, 2023). Many have resorted to utilizing detection tools for the presence of any AI input (Weber-Wulff *et al*., 2023). However, as demonstrated by Weber-Wulff et al. (2023), these tools carry a high percentage of error, which leads to an impairment in their reliability and in verifying authentic work. In response to these developments, teachers proceed to redesign assignment tasks and focus on the spoken component and in-class tasks (Perkins *et al*., 2024)

2.4. Challenges of Academic Assessment and Accreditation in the Era of Artificial Intelligence

The rapid spread of generative artificial intelligence technologies poses modern challenges to academic integrity, particularly its ability to produce texts that match or even surpass human production, outperforming detection tools (Weber-Wulff *et al*., 2023). The situation has prompted researchers and stakeholders to pay close attention to this issue and focus more on the oral component (Perkins *et al*., 2024). In conclusion, the development of AI-assisted writing has significantly changed the field of teaching and evaluating English. AI technologies have improved literary organization, fluency, and lexical sophistication, but they have also made it harder to distinguish between human and machine authorship, casting doubt on conventional ideas of authenticity and originality. When human and AI contributions combine, they raise significant challenges in maintaining academic integrity and assessing writing quality. This duality poses important challenges regarding the maintenance of academic integrity and the assessment of

writing quality. Teachers must reevaluate assessment literacy in light of these changes, especially their capacity to evaluate writing performance, create reliable assessment rubrics, and render morally sound decisions in AI-mediated settings. An emerging and understudied field in applied linguistics and language assessment research is the interaction between instructors' assessment beliefs, their evaluation methods, and their ethical issues regarding AI use.

3. Methodology

This study adopted the mixed approach method: quantitative and qualitative data research in the process of data collection. Questionnaires will be used, but with a number of lecturers. Interviews were conducted with lecturers who participated in the questionnaire.

3.1. Participants and sampling

The participants of this study were lecturers from the University of Benghazi language faculties. A total of 30 lecturers participated in the questionnaire; 16 (54%) were female and 14 (46%) were male. The interview participants consisted of nine individuals, with six males and three females. All of the participants in this study are proficient in English as a second language.

3.2. Data Collection Instruments and Procedures

Data were gathered from the English Language Department of the University of Benghazi. Permissions were obtained from university administrators.

The main instruments used to gather data were:

1. 30 lecturers were given questionnaires to complete to collect quantitative information about their perception of generative AI in relation to language assessment.
2. Structured Interviews: eight lecturers were interviewed to get detailed, qualitative information. The interviews were conducted in English.

3.3. Data Analysis

The data was examined to ascertain the instructors' perceptions and attitudes regarding generative AI in relation to language assessment.

Quantitative Analysis (Questionnaires): Microsoft Excel statistical analysis was used to examine the answers to the 30 questionnaires. To identify trends and patterns in the teachers' responses, the analysis concentrated on calculating frequencies and percentages, when appropriate.

Qualitative Analysis (Interviews): The interviews were carefully written and audio-recorded before being reviewed. The researcher examined the interviewees' answers to identify significant similarities and differences in their attitudes, experiences, and behaviors.

3.4. Validation of Instruments

There were two stages to the validation procedure for the questionnaire. It was first examined by four specialists in EFL instruction and educational research, whose comments resulted in modest phrasing changes and the elimination of unclear passages. To evaluate item clarity and internal consistency, a pilot test was conducted with three teachers who were not part of the main sample. The same group of experts examined the interview protocol to make sure it was clear and in line with the goals of the study.

3.5. Fieldwork difficulties

The researcher faced field difficulties in distributing questionnaires on paper due to the long distance and the challenge of continuous travel between scattered colleges. The differing schedules and lecture timings, along with the varying free times among faculty members, prevented direct access and personal meetings with them. Therefore, Google Forms was used to design the electronic questionnaire and facilitate quick access for participants.

3.6. Limitations

1. The study's spatial scope was restricted to the colleges of the University of Benghazi, which are located in three distinct cities.

2. The study was insufficient for absolute generalization due to the limited and small sample of faculty members.

3. The current results cannot be applied to other institutional contexts due to the small size of the research sample.

4. The comprehensiveness and accuracy of the results necessitate the completion of this phenomenon through broader future studies and larger samples.

4. Results

4.1. Quantitative Findings

4.1.1. Section 1: Perceptions of the Impact of Generative AI on Language Assessment

This section reports responses to the first six questionnaire items, which addressed participants' (N = 30) perceptions of generative AI's impact on writing assessment. Items were rated on a 5-point Likert scale (1 = strongly disagree, 5 = strongly agree).

Item 1: *Generative AI has fundamentally changed what it means to assess a student's writing ability.* Most participants agreed with this statement, with 13 (43.3%) selecting "agree" and 8 (26.7%) selecting "strongly agree," for a combined agreement rate of 70.0%. Five participants (16.7%) were neutral, while four (13.3%) disagreed or strongly disagreed (2 each). The item produced a mean of 3.77 (SD = 1.14), indicating an overall tendency toward agreement.

Item 2: *Generative AI threatens the validity of take-home writing assignments.* This item produced the strongest endorsement in the section. Eighteen participants (60.0%) strongly agreed and six (20.0%) agreed, yielding 80.0% combined agreement, the highest of any item. Only one

ISSN: 4556-3489
Arab Journal of Advanced Research and Development (AJARD)
المجلة العربية للبحوث المتقدمة والتنمية

participant (3.3%) was neutral, and five (16.7%) disagreed or strongly disagreed (3 and 2, respectively). The mean was 4.17 (SD = 1.29), the highest recorded across the section.

Item 3: *Generative AI threatens the validity of in-class, unsupervised assessments.* Responses to this item were more divided than for take-home assessments. Ten participants (34.5%) strongly agreed, and four (13.8%) agreed, for a combined agreement. Eight participants (27.6%) were neutral. Seven participants (24.1%) disagreed or strongly disagreed, with six disagreeing and one strongly disagreeing. The mean was 3.59 (SD = 1.21).

Item 4: *Generative AI makes it harder to distinguish genuine student ability from AI-assisted output.* Agreement on this item was more evenly split than on others. Eight participants (26.7%) strongly agreed, and eight participants (26.7%) agreed, resulting in a combined agreement rate of 53.3%, while four participants (13.3%) were neutral. Ten participants (33.3%) disagreed or strongly disagreed (7 and 3, respectively), the highest disagreement rate in the section. The mean was 3.37 (SD = 1.38), the lowest of all items, reflecting this divided opinion.

Item 5: *Generative AI has a greater impact on writing assessment than on speaking, listening, or reading assessment.* Nine participants (30.0%) strongly agreed and ten (33.3%) agreed, for a combined agreement. Seven participants (23.3%) were neutral, the second-highest neutrality rate in the section, while four (13.3%) disagreed or strongly disagreed (1 and 3, respectively). The mean was 3.70 (SD = 1.24).

Item 6: *Institutional assessment policies have failed to adapt to the influence of generative AI.* Fourteen participants (46.7%) strongly agreed and nine (30.0%) agreed, for a combined agreement. Five participants (16.7%) were neutral. Most participants agreed, and only two (6.7%) strongly disagreed, the lowest disagreement rate of any item. The mean was 4.10 (SD = 1.12).

Table 1: Frequencies, means, and standard deviations for items 1-6

No.	Item	M	SD	Agree/SA%	Neutral%	Disagree/SD%
1	Changed what it means to assess writing	3.77	1.14	70.0%	16.7%	13.3%
2	Threatens validity of take-home assignments	4.17	1.29	80.0%	3.3%	16.7%
3	Threatens validity of in-class/unsupervised	3.59	1.21	48.3%	27.6%	24.1%
4	Harder to distinguish genuine vs. AI-assisted output	3.37	1.38	53.3%	13.3%	33.3%
5	Greater impact on writing than other skills	3.70	1.24	63.3%	23.3%	13.3%
6	Institutional policies have not kept pace	4.10	1.12	76.7%	16.7%	6.7%

Note. SA = strongly agree; SD (in mean columns) = standard deviation; SD (in percentage column) = strongly disagree.

4.1.2. Section 2: Extent of Influence on Testing and Assessment Practices

0096893649662 | info@ajard.org | www.ajard.org

ISSN: 4556-3489
Arab Journal of Advanced Research and Development (AJARD)
المجلة العربية للبحوث المتقدمة والتنمية
Vol. 2, No. (8) Sep. 2026

This section reports responses to five items assessing the degree to which generative AI has influenced specific aspects of participants' (N = 30) assessment practices, rated on a 5-point scale (1 = not at all, 2 = slightly, 3 = moderately, 4 = greatly, 5 = to a very great extent).

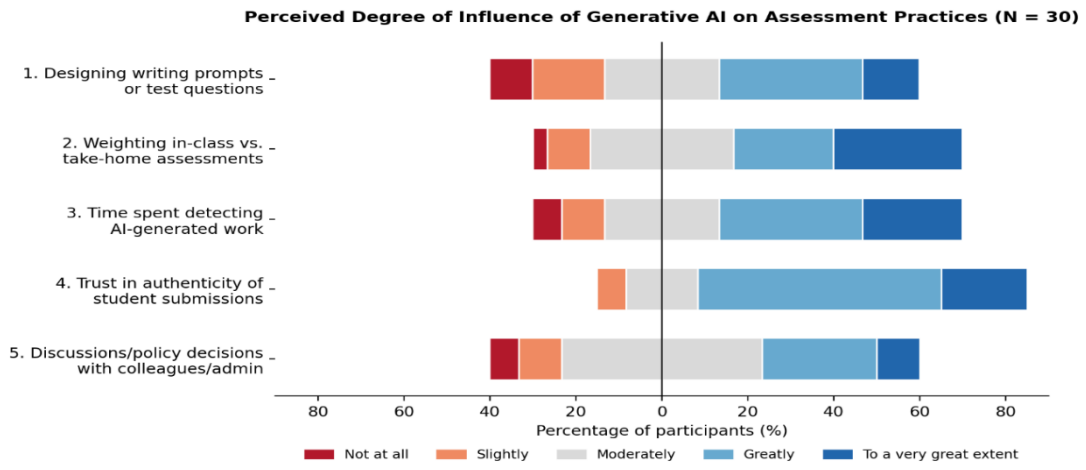
Item 1: *It influences the way you design writing prompts or test questions.* Influence on prompt design was moderate overall. Four participants (13.3%) reported influence to a very great extent and ten (33.3%) reported great influence, for a combined 46.7% reporting substantial influence. Eight participants (26.7%) reported moderate influence, while eight (26.7%) reported slight or no influence (5 and 3, respectively). The mean was 3.23 (SD = 1.19).

Item 2: *It influences the weighting you assign to in-class vs. take-home assessments.* Nine participants (30.0%) reported influence to a very great extent, and seven (23.3%) reported great influence, for a combined total of 53.3%. Ten participants (33.3%) reported moderate influence, and four (13.3%) reported slight or no influence (3 and 1, respectively). The mean was 3.67 (SD = 1.12).

Item 3: *It influences how much time you spend detecting possible AI-generated work.* Seven participants (23.3%) reported a very great extent of influence, and ten participants (33.3%) reported a great influence, resulting in a combined total of 56.7%, which is the second-highest substantial-influence rate in this section. Eight participants (26.7%) reported moderate influence, and five (16.7%) reported slight or no influence (3 and 2, respectively). The mean was 3.57 (SD = 1.17).

Item 4: *It influences your trust in the authenticity of student submissions overall.* This item showed the strongest and most concentrated influence in the section. Six participants (20.0%) reported influence to a very great extent, and seventeen (56.7%) reported great influence; for 76.7% combined, this was the highest rate recorded. Five participants (16.7%) reported moderate influence, two (6.7%) reported slight influence, and no participant reported no influence at all. The mean was 3.90 (SD = 0.80), also the lowest variability of any item, indicating unusually strong consensus.

Item 5: *It influences discussions or policy decisions with colleagues/administration about assessment.* This item showed the weakest reported influence. Three participants (10.0%)



0096893649662 | info@ajard.org | www.ajard.org

reported influence to a very great extent, and eight (26.7%) reported great influence, for a combined total of only 36.7%. Fourteen participants (46.7%), the largest single group for any response option in this section, reported moderate influence, and five (16.7%) reported slight or no influence (3 and 2, respectively). The mean was 3.23 (SD1.01).

Figure1: *Perceived degree of influence of generative AI on assessment practices.*

Finally, when asked to describe the overall influence of generative AI on their assessment practices, over half of the respondents reported a major influence or complete transformation (n = 17). A further (n = 8) participants noted a moderate influence, while a minority indicated either a minor influence (n = 3) or no influence at all (n = 2).

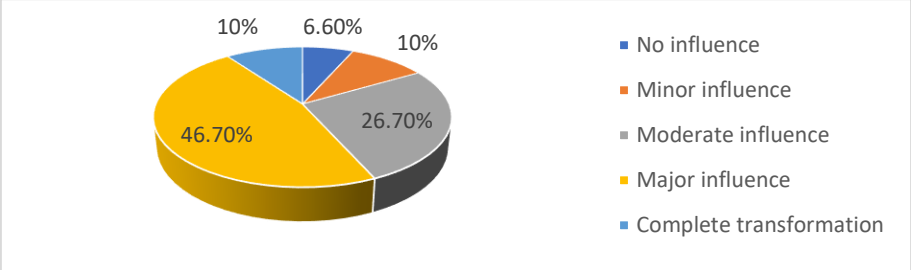


Figure 2: *Overall influence of generative AI on assessment practices.*

4.1.3. Section 3: Changes Made to Assessment Methods in Response to Generative AI

This section reports N = 30 participants' self-assessed confidence in adapting to generative AI, perceived need for institutional support, and intentions for further change. The data were collected using a 5-point Likert scale (1 = strongly disagree, 5 = strongly agree). It also details the multiple-choice item-determined changes participants have made to their assessment practices in response to generative AI.

Item 1: *I can adapt my evaluations to generative AI. Responses were neutral. Five participants (16.7%) strongly agreed and eight (26.7%) agreed. Twelve participants (40.0%) had the highest neutrality rate in all sections. Five participants (16.7%) disagreed or strongly disagreed, four and one, respectively. The lowest mean and variability of the three items was 3.40 (SD = 1.04). This indicates lukewarm and uniform confidence.*

Item 2: *I need institutional support and training to redesign assessments for AI. It generated the section's strongest and most polarized response. The highest agreement rate in this section was 70.0%, achieved by fourteen participants (46.7%) who strongly agreed and seven (23.3%) who agreed. Six (20.0%) disagreed or strongly disagreed, while three (10.0%) were neutral. The highest and most variable mean was 3.87 (SD = 1.38). Participants were divided between those who strongly desired additional support and those who did not feel the need for it.*

ISSN: 4556-3489
Arab Journal of Advanced Research and Development (AJARD)
المجلة العربية للبحوث المتقدمة والتنمية
Vol. 2, No. (8) Sep. 2026

Item 3: *I will soon change my evaluation methods. Eight participants (26.7%) strongly agreed, and twelve (40.0%) agreed. Six (20%) disagreed or strongly disagreed, while four (13.3%) were neutral (4 and 2). A mean of 3.67 and 1.21 standard deviation were found.*

Table 2. Frequencies, means, and standard deviations for confidence, support, and intention items

No.	Item	M	SD	Agree/SA%	Neutral%	Disagree/SD%
1	Confidence adapting assessments to AI	3.40	1.04	43.3%	40.0%	16.7%
2	Wants more institutional support/training	3.87	1.38	70.0%	10.0%	20.0%
3	Plans further changes to assessment methods	3.67	1.21	66.7%	13.3%	20.0%

Note. SA = strongly agree; SD (in mean columns) = standard deviation; SD (in percentage column) = strongly disagree. N = 30 for all items.

4.2. Changes Already Made in Response to Generative AI

Participants reported making multiple concurrent changes to their assessment practices, and all participants reported making changes (100%). The most common adaptation was increased use of in-class, handwritten, or supervised assessments, selected by 20 participants (66.7%), by far the most widely adopted response. Reducing the weight of take-home assignments in final grades followed at 12 participants (40.0%), and explicitly teaching students about appropriate/inappropriate AI use at 10 participants (33.3%). Adding oral defenses or viva-style follow-ups to written work was reported by 9 participants (30.0%), while incorporating AI-detection software was the least common change, at 5 participants (16.7%). Since respondents could select multiple options, percentages do not sum to 100%; the pattern suggests participants have favored procedural and pedagogical adaptations, changing assessment format, and teaching AI norms over technological detection as their primary response.

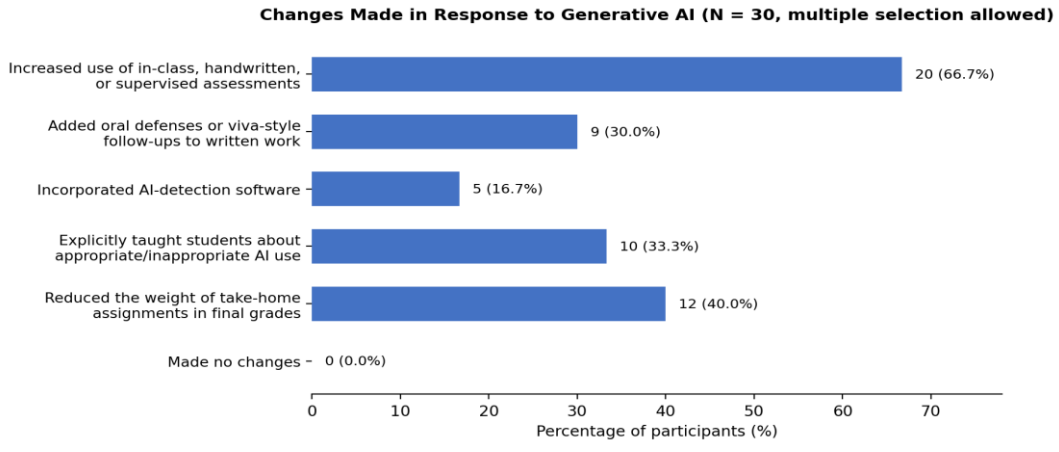


Figure 3: *Changes made in response to generative AI (multiple selection allowed).*

4.3. Interview Analysis (Qualitative Findings)

0096893649662 | info@ajard.org | www.ajard.org

The researcher applied the structured interview method when conducting his interviews with the respected lecturers; all the participants were asked the same set of questions. The participants were asked seven questions that were predetermined and written in advance to maintain consistency and compare data between participants. Questions were written in specific detail; the sequence of questions was the same across the interviews. The questions were designed to address the research paper's objectives. The questions were about generative AI's impact on testing and assessment, their views on technology, their future plans, and changes to their assessment methods.

Table 3: Interviewee information

No.	Participants		Gender	Age	Years of experience
1	Interviewee #1		Female	30	3
2	Interviewee #2		Male	33	7
3	Interviewee #3		Male	34	7
4	Interviewee #4		Male	44	15
5	Interviewee #5		Male	50	18
6	Interviewee #6		Female	45	11
7	Interviewee #7		Female	40	10
8	Interviewee #8		Male	37	9

The analysis of the interviews data has been organized into seven themes:

4.3.1. Theme 1: A Shift from Product-Based to Process- and Competence-Based Assessment

Across nearly all interviews, participants described a fundamental shift in what they believe testing should assess, moving away from evaluating a finished written product toward evaluating the underlying competence and thinking process that produced it. Interviewee 1 explained that a polished essay *"doesn't necessarily mean they can actually write or communicate at that level themselves*," and that assessment *"needs to measure more than the final product,"* including "*the student's own understanding, their ability to explain their ideas, and also how they use English spontaneously"* Interviewee 4 articulated a similar realization, stating that "*traditional homework can't accurately measure a student's true language skills anymore"* and that assessment *"must happen inside the classroom" to allow direct observation of "real abilities".* Interviewee 6 extended this shift toward metacognitive and reflective dimensions of language use, noting that assessment "*should also examine how students plan, revise, and justify their language choices"*.

0096893649662 | info@ajard.org | www.ajard.org

ISSN: 4556-3489
Arab Journal of Advanced Research and Development (AJARD)
المجلة العربية للبحوث المتقدمة والتنمية

Interviewee 7 framed the change most explicitly around critical thinking, arguing that "*we must test how students think, not just how they write,"* with an emphasis on students' capacity to *"build an argument, organize ideas, and question information"* and to demonstrate *"their real, human voice"*. Interviewee 8 reached a comparable conclusion, remarking that *"if an algorithm can pass the class for you, the test was pointless anyway"*. Collectively, these replies indicate a unified perspective among participants that legitimate evaluation in the generative AI era should focus on process, rationale, and impromptu execution rather than only the finished product presented.

4.3.2. Theme 2: Generative AI as a Dual-Natured Force, Threat and Opportunity

All eight interviews consistently portrayed generative AI as equally harmful and potentially advantageous for language assessment, rather than as a purely negative development. Interviewee 1 bluntly stated, "*I would say both*." She elucidated that AI is a threat when students "*produce assignments... without actually doing the language work themselves*," but it becomes valuable when integrated pedagogically. For example, students could *"write a first draft themselves, then use AI to obtain language suggestions, and afterward explain which suggestions they accepted and why.* "Interviewee 3 also identified an opportunity in the test-design process, describing AI as beneficial for "*creating an excellent exam paper*." However, she framed the threat as centered on "home-based assignment[s]." Interviewee 8 provided a specific example of transforming the threat into an assessment opportunity, describing a transition to evaluating "*their capacity to discuss, explain, and respond to inquiries in real time*" in order to make it more difficult to conceal the use of AI. On the other hand, Interviewee 5 relied more heavily on the threat framing, defining generative AI as a problem "*unless it is governed or harnessed*" and cautioning that an overreliance on such tools poses "*a danger to human intellect in the long run*." This discrepancy suggests that, despite the fact that the majority of participants maintained a balanced, dual-natured perspective, a small number were more apprehensive about the potential for AI to impede the development of authentic language in the absence of robust institutional constraints.

4.3.3. Theme 3: Take-Home Essays and Written Assignments as the Most Affected Assessment Type

Every participant, without exception, identified take-home essays and written assignments as the assessment format most disrupted by generative AI, typically contrasted with in-class and oral tasks, which they viewed as comparatively resistant to AI misuse because they are directly observed. Interviewee 1 noted that with take-home writing *"it is very difficult to know how much of the final text actually belongs to them,"* pointing to a telling pattern in which a student's usual "*repeated grammatical mistakes or a relatively simple vocabulary"* would suddenly give way to *"very advanced vocabulary and complex sentence structures"*. Interviewee 2 offered a more compact version of the same distinction, stating that in-class exams and oral tasks *"are controllable... because they are being done in front of you,"* unlike essays. Interviewee 5 described the same discrepancy between in-class and take-home performance, observing that students who *"get stuck"* and produce writing full of *"grammar mistakes, unrelated ideas*" in class will nonetheless "*turn up with a wonderful essay that could have been written by a[n] educated native speaker"* when the task is completed at home. Interviewee 7 summarized the broader consequence of this pattern, concluding that "traditional homework has lost its validity" as an assessment tool.

0096893649662 | info@ajard.org | www.ajard.org

ISSN: 4556-3489
Arab Journal of Advanced Research and Development (AJARD)
المجلة العربية للبحوث المتقدمة والتنمية

4.3.4. Theme 4: Detection Practices and the Limits of AI-Detection Software

Many participants specifically warned against using AI-detection software (such as Turnitin's AI indicators) as independent proof of misbehavior, and several mentioned specific instances in which suspected AI use led to a change in assessment procedure. Interviewee 1 described a case in which unusually polished writing, combined with a Turnitin AI indicator, led not to an automatic penalty but to a conversation with the student, an approach that *"gave me a better idea of whether the student really understood what they had submitted"*; the experience taught the interviewee that "*the AI report can raise a concern, but the lecturer's professional judgement and direct evidence of the student's ability are also very important"*. This same participant later generalized the point, arguing that *"a Turnitin AI indicator... should not automatically be treated as proof that a student has used AI"* and that lecturers must instead "consider the whole picture" by comparing current and prior work. Interviewee 6 described a similar corroborating strategy, changing an assessment procedure after noticing *"unusually similar vocabulary and sentence structures"* across submissions by requiring in-class drafting plus a reflective explanation of the revision process. Interviewee 2 and Interviewee 5 both recounted instances of oral follow-up that revealed a discrepancy between their submitted responses and their actual comprehension: Interviewee 2 recalled that students who had "*answered them with AI tools and memorized the answers*" "*got full marks*" on a subsequent exam using the same questions. Interviewee 5 recounted asking students about their reading-comprehension answers orally and finding "*they had no idea about their answer*," ultimately failing them for not understanding the task they had submitted. In a similar vocabulary-task incident, Interviewee 4 recounted that "*many students submitted sentences that were nearly identica*l," which resulted in a permanent relocation of the activity to class time.

4.3.5. Theme 5: Toward Oral, In-Class, and Portfolio-Based Assessment

While consistently identifying the same structural barriers: time, class size, and unclear institutional policy, participants converged on a set of proposed reforms that included more oral assessment, in-class supervised writing, and portfolio or process documentation, limited by Policy and Workload. While acknowledging that "the primary issue is time" in cases of large class sizes, Interviewee 1 suggested evaluating "different stages: the outline, the first draft, the revision, and perhaps a brief oral discussion." Interviewees 4 and 7 emphasize the importance of oral exams, as they require students to "think and speak spontaneously without AI help. "Nevertheless, they both acknowledged the same constraint: Interviewee 7 asserted that "it is physically impossible to monitor the group dynamics and poster drafts of 50 or 60 students without institutional support or lighter teaching loads." Beyond burden, "the time of classes and nature of the course" restrict what is possible, according to Interviewee 3. Interviewee 8 also mentioned "the lack of clear guidance on how to assess students fairly when AI is involved." Interviewee 5 adopted a more stringent approach, advocating for examinations to be "carefully supervised and guarded against AI users "and for a complete elimination of take-home assignments.

0096893649662 | info@ajard.org | www.ajard.org

4.3.6. Theme 6: Envisioning Future Assessment as Continuous, Multimodal, and AI-Literate

When asked to look ahead, most participants described a hybrid future in which assessment becomes more continuous and multimodal rather than either banning AI or ignoring it. Interviewee 1 predicted a system combining "*a portfolio, classroom writing, oral presentations, short interviews, reflective tasks, and supervised writing,* "paired with explicit instruction in *"how to use AI responsibly"*. Interviewee 6 used similar language, envisioning assessment that is *"more continuous, multimodal, and process-oriented,"* combining *"controlled in-class tasks with carefully supervised use of AI for brainstorming, feedback, and revision,"* with students assessed on their ability *"to evaluate AI-generated language critically rather than simply accept it"*. Interviewee 7 described a comparable *"hybrid, two-zone"* model, pairing "*highly secure, in-person tasks"* with *"collaborative, open-AI projects"* graded on critique and oral defense of AI-assisted work. By contrast, Interviewee 3 and Interviewee 5 leaned toward a more restrictive vision, with Interviewee 3 calling for assessment to *"totally rely on strategies... that would make for authentic ability testing"* with no opportunity to use generative AI, and Interviewee 2 arguing that policymakers should *"force an anti-AI tool to control this"*. This split suggests an emerging tension in the field between an integrationist model, which teaches supervised and critical AI use, and a restrictive model, which seeks to exclude AI from assessment contexts entirely.

4.3.7. Theme 7: Overlooked Issues, Detector Overreliance, Equity, and the Pace of Change

In response to the final, open-ended question, participants raised several concerns not directly solicited by earlier questions, most notably equity of access to AI tools and the danger of overreliance on detection software. Interviewee 7 argued that *"advanced AI tools that provide the best academic feedback require paid subscriptions"* and warned that without free institutional access, "wealthier students will have a massive, unfair advantage". Interviewee 6 expressed a similar concern regarding "assessment fairness," observing that "students have unequal access to AI tools and varying levels of digital literacy" Interviewee 1 revisited the theme of detector overreliance, determining that the primary inquiry for the field is "how do we design assessment so that it actually demonstrates the student's capabilities?" This is "probably the most significant transformation that generative AI has introduced to language assessment." Interviewee 8 concluded by reflecting on the pace of change, stating that "we are all still learning how to address these challenges, teachers and students alike," following a description of a student who was unable to articulate AI-assisted edits to their writing. Interviewee 5 expressed a more apprehensive assessment of the technology's trajectory, stating that the distinction between AI-generated and human-produced text or speech is now "impossible" as generative AI "becomes more human-like in its behavior."

5. Discussion

The results suggest that lecturers are cognizant of the impact of generative AI on student evaluations. The validity of take-home assignments was a source of concern for the majority of participants, while opinions regarding in-class assessments were more divided. This indicates that

ISSN: 4556-3489
Arab Journal of Advanced Research and Development (AJARD)
المجلة العربية للبحوث المتقدمة والتنمية

classroom management continues to be a significant factor in the evaluation of students' work. The challenge of distinguishing between authentic student work and AI-assisted work was another significant concern that elicited some of the most polarized responses from participants. The findings also suggest that generative AI has a significant influence on writing skills.

In comparison to supervised, in-class writing tasks, the majority of participants believed that unsupervised writing assignments are particularly susceptible to AI-assisted work. The results also underscore the necessity of implementing institutional policies and support to enhance assessment practices and enable institutions to keep up with the rapid advancement of AI technologies. However, a small number of participants indicated that generative AI has had a limited impact on their assessment practices. The data indicated that generative AI had a substantial impact on the trust that lecturers had in the authenticity of students' submitted work, as evidenced by their evaluations of in-class versus take-home assignments. This implies that generative AI has altered a more fundamental and widely shared phenomenon: a baseline erosion of trust in the authenticity of submitted work, irrespective of the concrete changes individual lecturers have or have not made to their assessment practices.

The analysis also demonstrated that lecturers had varying perspectives on the impact of generative AI. the impact of generative AI The analysis also revealed a diverse range of perspectives among lecturers regarding the impact of generative AI. While a second group perceived its influence as moderate, the dominant position among lecturers indicated a significant impact on assessment practices. Conversely, a small number of lecturers reported that generative AI had a negligible or no impact on their assessment practices. The results indicated that a substantial number of participants intended to modify their assessment methods, while a comparable number had already implemented at least one concrete change. Additionally, the results indicated a disparity in the degree of confidence that lecturers had in their capacity to modify assessment practices in response to the implementation of generative AI.

Simultaneously, there was a strong desire for institutional support and training, particularly among lecturers who felt unsupported in adapting their assessment practices to the challenges posed by generative AI. A consistent image was discerned throughout the eight interviews. The majority of participants concurred that generative AI has compromised the validity of take-home written assessments while also recognizing AI as a dual-natured tool that presents both challenges and potential advantages. They indicated a preference for in-class, process-based, and oral assessments, and participants also raised recurring concerns about equitable access to AI tools, supplemented by reflective or explanatory components. Participants also exercised caution when considering AI-detection software as conclusive evidence, opting to verify suspected AI use through direct questioning and students' prior performance. Nevertheless, the interviews underscored the logistical and institutional obstacles to assessment reform, including the absence of clear institutional policies, insufficient time, and large class sizes. Participants also expressed ongoing concerns about equitable access to AI tools. Participants also expressed ongoing concerns about equitable access to AI tools. also raised.

6. Conclusion

In conclusion, the findings of this study, which are corroborated by the interview data, indicate that generative AI has significantly altered the way in which lecturers perceive the current assessment practices in the field. The widespread use of these tools has resulted in significant changes in the attitudes of lecturers toward homework and take-home projects that are completed without direct supervision. This has raised concerns about the legitimacy and authenticity of students' written work. In response to these apprehensions, lecturers have begun to modify certain

0096893649662 | info@ajard.org | www.ajard.org

ISSN: 4556-3489
Arab Journal of Advanced Research and Development (AJARD)
المجلة العربية للبحوث المتقدمة والتنمية

aspects of their evaluation procedures to facilitate the identification of whether students' work is indicative of their knowledge and skills. The results also indicate a heightened emphasis on critical thinking and assessment methods that provide students with the opportunity to elucidate and exhibit their comprehension. In addition, it is imperative that educational authorities and decision-makers offer lecturers explicit guidance and assistance as they adjust to the implementation of generative AI. It will be crucial to update assessment methods to preserve academic integrity and guarantee that assessments continue to accurately reflect students' learning.

7. Recommendations

Here is a number of proposed recommendations based on the key findings of the current study:

1- Conduct training workshops to assist lecturers in enhancing their abilities to effectively manage AI tools.
2- Revise the manner in which we conduct university grading to match the latest technology. Normal homework should be discontinued in favor of in-class assignments and step-by-step grading.
3- Request that the government and university leaders provide clear plans and assist departments in establishing effective regulations to address these issues.
4- Conduct additional research in this field to identify alternative solutions and remain informed about the rapid expansion of AI.
5- Emphasize the evaluation of each student's critical thinking and direct speaking (such as speaking tests and presentations) as significant components of student grading to ensure that the results are accurate and genuine.

References

Al-Oteawi, S. M. (2002). *The perceptions of teachers and library media specialists on the integration of information technology in secondary schools in Saudi Arabia* [Doctoral dissertation, Ohio University]. OhioLINK Electronic Theses and Dissertations Center.

Albion, P. R., & Ertmer, P. A. (2002). Beyond the foundations: The role of vision and belief in teachers' preparation for integration of technology. *TechTrends*, 46(5), 34–38.

Becker, H. J., Ravitz, J. L., & Wong, Y. T. (1999). *Apple Classrooms of Tomorrow–Today. Professional development and computer use: Teacher and school factors related to instructional computer use among elementary and secondary schoolteachers* (Report No. 4). Center for Research on Information Technology and Organizations, University of California, Irvine.

Chiu, T. K. (2023). The impact of Generative AI (ChatGPT) on teaching and learning: A study of student and teacher perceptions. *Computers and Education: Artificial Intelligence*, 4, Article 100147.

Cotton, D. R., Cotton, P. A., & Shipway, J. R. (2024). Chatting and cheating: Ensuring academic integrity in the era of ChatGPT. *Innovations in Education and Teaching International*, 61(2), 228–239.

Eaton, S. E. (2023). *Postplagiarism: Transdisciplinary essays on the ethical and educational implications of ChatGPT*. Springer.

Gardner, H., Kornhaber, M. L., & Wake, W. K. (1993). *Intelligence: Multiple perspectives*. Harcourt Brace College Publishers.

Hogg, M. A., & Vaughan, G. M. (2005). *Social psychology* (4th ed.). Prentice Hall.

Perkins, M., Roe, J., Postma, D., & McGaughran, J. (2024). Game of tokens: University student and lecturer perceptions of generative AI in higher education. *Computers and Education: Artificial Intelligence*, 6, Article 100226.

Saaid, M. (2010). Technology integration in education: Barriers and solutions. *Journal of Educational Technology Systems*, 38(2), 155–170.

Weber-Wulff, D., Anohina-Naumeca, A., Bjelobaba, S., Foltýnek, T., Guerrero-Dib, J., Hrek, O., ... & Zincke, C. (2023). Testing of detection tools for AI-generated text. *International Journal for Educational Integrity*, 19(1), Article 26.

Wozney, L., Venkatesh, V., & Abrami, P. C. (2006). Implementing computer technologies: Teachers' perceptions and practices. *Journal of Technology and Teacher Education*, 14(1), 173–207.

Zammit, S. (1992). Factors facilitating or hindering teachers' use of information technology in school. *Journal of Computer Assisted Learning*, 8(1), 57–66.