Trustworthiness of EFL Assessment of Learning in the Age of AI: Challenges and Solutions

Iman El-Nabawi Abdel Wahed Shaalan[1], Ayman Shaaban Khalifa Ahmad[2]

[1] Department of English, College of Science and Humanities, Prince Sattam Bin Abdulaziz University, KSA, Al-Kharj 11942, Kingdom of Saudi Arabia

[2] Faculty of Education for Boys, Al-Azhar University, Cairo, Egypt

Correspondence: Iman El-Nabawi Abdel Wahed Shaalan, College of Science & Humanities, Prince Sattam bin Abdulaziz University, Saudi Arabia. E-mail: i.shalaan@psau.edu.sa

Received: September 13, 2024 Accepted: November 8, 2024 Online Published: March 10, 2025

doi:10.5430/wjel.v15n5p1 URL: https://doi.org/10.5430/wjel.v15n5p1

Abstract

This study aimed to explore the assessment trustworthiness of English as a Foreign Language (EFL) in the Artificial Intelligence (AI) age by identifying the main challenges and proposing viable solutions. Employing a qualitative case study approach, the research investigated the experiences and perceptions of EFL instructors regarding the challenges and solutions. To meet such an end, the study sought, through semi-structured interviews, to gain insights from the study sample which comprised nine experienced EFL instructors selected based on their expertise in the field of EFL teaching and AI technology. The findings revealed numerous significant challenges, including the disadvantageous effect of AI tools on academic integrity, classwork engagement, reliance on technology, students' creativity, and current assessment metrics. Despite such challenges, the study portrayed some effective solutions, such as designing authentic assessment tools for assessing higher cognitive skills, adopting active learning strategies, developing training programs for EFL learners, implementing advanced AI content detectors, and updating traditional assessment methods. Based on the results, the study highlighted a dire need to reform conventional assessment practices to address integrity challenges posed by AI tools.

Keywords: trustworthiness, assessment, challenges, solutions

1. Introduction

Playing a pivotal role in promoting effective learning, assessment is superior to other classroom activities and the teaching-learning process elements (Amua-Sekyi, 2016). Objective assessment practices serve as a means for presenting a comprehensive picture of the learning progress to the stakeholders, including students, teachers, parents, supervisors, education planners and developers (Suskie, 2018). Trustworthy assessment practices ensure that deep learning, critical thinking, and fine-tuned learning outcomes are achieved in line with the specific requirements of the education quality standards and the labour market. Assessment results' outcomes are crucial for custom-tailoring the teaching-learning process to maximise such outcomes and encounter the challenges (Driscoll & Wood, 2023).

Assessment has a considerable role in promoting the efficiency and quality of the educational process (Al Ghamdi & Al-Ghamdi, 2021; Tosuncuoglu, 2018), and it also directs the efforts exerted for enhancing the teaching-learning situation such as curriculum development, adopting new teaching methods and adjusting the educational policies considering the new related variables (Abdulrazzaq & Abdellatif, 2023; Zhao, 2024). Another point to be considered is that realistic assessment practices empower teachers to provide feedback and personalise the teaching-learning process; consequently, the teaching-learning practices are also adapted in light of these assessment tools (Kushari & Septiadi, 2022).

Assessment in general and EFL assessment in particular vary to include various tasks such as writing tasks, presentations, and projects (Abd Elgalil et al., 2022; Gultom, 2016). Another item of interest is the new advancements in the technological field targeting the assessment to help teachers create tests, provide feedback, and track the examinees' progress (Chen, 2020; Shaalan & Ahmad, 2024). More importantly, the adaptive assessment systems and the collaborative assessment practices have become an integrated part of the modern educational context (Gusev & Armenski, 2014; Menard, 2021).

Principally, assessment encompasses three types: initial assessment (conducted at the beginning of the instruction process to assess the learners' background knowledge), formative assessment (performed during the instruction process regularly to ensure the acquisition of the target learning concepts and enhance the learning process), and summative assessment (performed at the end of the instruction process for measuring the overall acquisition of the target knowledge and skills) (Cheng et al., 2004; Wei, 2015). The summative assessment was proved effective in developing the students' performance and a proper indicator of the achievement of the instructional goals and proficiency of learning. Furthermore, it improves the retention and use of the acquired knowledge and skills (Marzano, 2010).

Likewise, assessment is a worthwhile means for ensuring equity via the norm-referenced criteria; subsequently, assessment bias directly

causes inequalities, resulting in future negative implications. Assessment bias means disparities that negatively impact the learners in the present and future (Romano, 2024). Fostering a learning environment that sponsors equality and transparency in assessment is one of the main critical features of every successful teaching endeavour (Krzykowski & Kinser, 2014).

The employment of AI tools to assist EFL students in accomplishing their work is one of the educational dilemmas which have recently attracted the attention of educationalists and stakeholders in the education context (Abd El-Magid, 2024; Javed, 2024). In other words, EFL learners can utilise AI generative tools such as ChatGPT, Copilot, Gemini and other countless tools to complete homework, answer difficult questions, or generate well-developed content in different disciplines with higher levels of accuracy (Smolansky, 2023). Such practices negatively affect the academic integrity of the assessment process, which impacts the fairness of the assessment process because of using unauthentic materials generated by AI tools. Thus, employing AI in the teaching-learning process has ignited heated debate regarding its benefits versus its negative consequences (Smolansky, 2023; Zohny et al., 2023).

Although AI generative tools are to be implemented in the teaching-learning process to enhance critical thinking, creativity, and several other higher mental skills, a loss of creativity was uncovered among the students using generative AI for accomplishing their class tasks. Such acts necessitate a balance regarding the reliance on AI generative tools (Habib et al., 2024). More importantly, the content generated by AI tools is not attributed to their founders, and the accuracy of the results is always doubted. Such practices necessitate urgent actions and precautions to ensure transparency and integrity (Hristov, 2016). Another main negative consequence of the blind employment of AI tools in the teaching-learning process is the dependence on AI tools for validating the pieces of information and receiving feedback concerning the debatable issues that might promote what can be called the "tyranny of the majority" and the disregard of the minority needs (Lindsay et al., 2023). More critically, overreliance on AI tools for generating content and doing school tasks has caused several psychological problems. Students might feel apprehensive and anxious because of their uncertainty concerning responses submitted to their teachers. Furthermore, overdependence on AI tools might impact learners' autonomy (Han et al., 2023; Xie et al., 2022).

Subsequently, assessment reform practices are urgently needed to consider the significance of concentrating on skills mastery and knowledge acquisition. The traditional assessment with its current format is easily compromised by AI tools, which necessitates attention to address such concerns. A substantial portion of academic work comprising assignments, essays, and projects is done by the students without supervision from the instructors; consequently, the learner's responsibility is a critical issue that needs to be considered by the reformers of the assessment process in the AI age. In other words, the students themselves hold a substantial portion of responsibility regarding the employment of AI tools for undertaking schoolwork (Moya & Eaton, 2023; Smolansky, 2023).

AI tools, with the recent fast updates and enhancements, have become powerfully supported and backed by massive stores of knowledge that allow them to manage school tasks and home assignments easily in no time. In writing, for instance, AI tools can easily produce pieces of writing with higher levels of accuracy and precision; the dependence on AI tools in executing such home assignments can be a handy option for some students (Atlas, 2023). Due to AI tools' inability, till the current time, to cite the references properly as they merely generate content, the concept of "ghost authorship" is dominating in such pieces (Hristov, 2016). On the other hand, the recent developments of AI tools have enabled such tools, with higher levels of accuracy, to answer reading comprehension questions regardless of the level of the questions: literal, inferential, and evaluative. With no need for external resources, AI tools can answer many questions, rephrase the answers, and draw conclusions (Yu, 2022). Developing feasible strategies to detect and prevent the unethical usage of AI is necessary in the current era to guarantee the fairness of assessment (Meek, 2016).

Constructive Learning Theory posits that meaningful engagement in the learning process to construct knowledge effectively is the cornerstone of learning, rather than the passive reception of theoretical information. Designing authentic assessments, evaluating the candidates' background knowledge and skills targeted and focusing on critical thinking and problem-solving might empower EFL learners to apply language skills in practical contexts (Jonassen, 1999). Thus, active learning strategies and authentic assessments are feasible solutions for the threats posed by the unethical adoption of AI. On the other hand, emanating from Bloom's Taxonomy, the hierarchical model of learning objectives categorisation, assessments ought to marginalise basic skills and focus on higher-order thinking skills (Krathwohl, 2002; Musa, 2023). Thus, deeper intellectual engagement can be fostered by addressing the upper tiers of Bloom's Taxonomy to lessen the dependence on AI activities.

Moreover, the theory of Technological Determination highlights that overreliance on technology might lead to societal and cultural changes, which might in turn impact educational practices. The abundance of AI tools might influence the integrity of assessment and teaching strategies necessitating the development of new strategies and techniques to guarantee the trustworthiness of assessments (Chandler, 1995). Furthermore, the proliferation of AI tools threatens assessment validity which is the cornerstone of assessment design. On the other hand, the validity theory, articulated by Messick (1989), focuses on the fairness of assessment and necessitates the need for innovative assessment practices to align with the contemporary challenges posed by AI.

To conclude, the employment of technology enhances the quality and efficiency of learning outcomes (Ahmad et al., 2021). The advent of AI indicates a new era of transformation in the educational landscape with the countless services it provides for students, teachers, and education policymakers. However, such promises are hindered by numerous threats that might contaminate the efficient implementation of AI tools in the teaching and learning process. Trustworthiness, as an inevitable component of every successful EFL assessment, has been threatened by the dependence on AI tools for generating answers to questions. Such a critical threat necessitates thorough exploration to take the necessary measures and precautions to guarantee a fair and valid assessment that can be used as a foundation for

making future decisions.

The adoption of AI in the EFL teaching-learning process has numerous significant implications. Jeopardised by AI technologies, the trustworthiness of the EFL assessment is one of the critical and urgent aspects that must be met to guarantee fair, valid, and reliable assessment. The present research sought to explore the challenges and solutions of the EFL assessment trustworthiness by answering the following questions:

- What challenges do EFL instructors encounter that may affect the trustworthiness of assessments?
- What solutions can address the challenges faced by the EFL instructors for ensuring trustworthy assessments?

2. Literature Review

A plethora of studies were conducted addressing the students' dependence on AI tools. For instance, Playfoot et al. (2024) investigated the students' willingness to use ChatGPT to complete their writing assignments. It was revealed that 32% of the study participants (totalling 160) showed willingness to use ChatGPT, and 15% currently use it. Although the study showed that personality features, study skills, and academic performance factors could not predict the future use of ChatGPT, it was revealed that the students' use of ChatGPT increases when the risk of being discovered is low and the punishment is light. Ibrahim et al. (2023) portrayed the idea of using ChatGPT as an AI tool accessible by most students, and they can use it easily for accomplishing their homework. However, it was revealed that there is a serious risk of plagiarism in the tasks posed by ChatGPT on two introductory advanced courses. The scores attained by students who used ChatGPT to accomplish their tasks were near finals. More critically, the students' intentional committing of some typos or punctuation mistakes makes the tasks submitted undiscoverable by AI content detectors.

Ghimire and Edwards (2024) reported that students use AI tools in the field of computer programming to accomplish their assignments and draft their research. The study targeted exploring the students' use of AI tools in an introductory programming course. The results revealed that students used AI tools excessively for solving problems, answering questions, and debugging; however, they did not use such tools to generate codes directly. The students also expressed their satisfaction and appreciated AI tools as they helped them enhance their learning outcomes positively. Rossell ó-Geli (2022) conducted a study that raised concerns about the future of teaching, learning, and assessment considering the employment of AI tools. The students were principally using AI tools to review essays and reports, find gaps and suggest modifications. The study concluded that the challenges of "how to assess?" and "how to teach?" require urgent solutions from the educationalists and stakeholders.

Chan (2024) explored the moral discomfort which students might suffer when using AI tools to accomplish their tasks and assignments. The AI Guilt as conceived by the study was investigated by using qualitative methodologies considering the social and psychological impact and its implications on the teaching-learning process. A plethora of findings were quoted, such as lack of work authenticity, fear of negative judgment, and self-efficacy. The study recommended refining the academic integrity measures and reconsidering the assessment process in general. The study, furthermore, suggested providing learners with ethical guidelines to raise the students' awareness about the proper usage of AI tools for harnessing the learning outcomes as well as decreasing AI guilt.

Al-Jarf (2024) explored university staff members' reactions to the students' employment of AI tools for completing their tasks. The study developed a survey, which was completed by forty-five instructors. The results of the study revealed that 98% of the instructors rejected the tasks that were generated by AI tools, and they asked their students to rewrite them. The study suggested acting in terms of updating the university policies regarding the employment of AI tools as well as making use of reliable AI plagiarism detection tools. Yeo (2023) tackled the writing skill in the age of AI where AI tools can entirely develop reports, articles, or essays without the need for adding a single idea or composing a single sentence. The study portrayed the problem of authorship and academic integrity. The study recommended an urgent search for creative ways to overcome the unethical use of AI by students to accomplish their tasks and homework.

Based on reviewing available literature, it was revealed that a sizeable portion of students willingly use AI tools for accomplishing their tasks, which varied to include writing essays, writing reports, or solving programming issues. However, the students' usage of AI tools is aligned with some psychological burden: AI Guilt (afraid of being discovered, and it is influenced by the degree of punishment that they might face). AI tool usage poses a serious risk of plagiarising the content, and some students try to trick the AI detectors by committing some grammatical mistakes or typos. Most studies call for urgent procedures to be taken to overcome such problems and ensure the trustworthiness of the assignments accomplished by the students.

Reflecting on this, the present research might contribute to the body of literature by addressing a critical gap, namely understanding the challenges that contaminate the trustworthiness of assessment and providing practical solutions for EFL instructors. The results attained from the research might depict a comprehensive outline of the status-quo of AI tools used by students and the implications of employing AI in the teaching and learning process for policymakers to promote educational practices and maximise learning outcomes. The present research aimed to pave the way for future direction for researchers to empirically examine the solutions provided by EFL instructors for overcoming the students' overreliance on AI tools for accomplishing their tasks. The present research sought to propose authentic and innovative assessment practices which encourage critical thinking and creativity and overcome the reliance on AI tools. Finally, the present study sought to maximise the instructors' perspectives regarding their perceptions about trustworthy assessment practices.

3. Method

Ontologically, the present research is grounded in interpretivism underscoring that reality is subjective, not objective, and it is constructed within a social context and shaped by the perceptions and experiences of individuals. In other words, the present research sought to draw a comprehensive picture of the challenges posed by the AI impact on assessment as well as propose some solutions as indicated by the EFL instructors. The interpretivism methodology enables profound understanding, encompassing multiple realities of the EFL instructors' perceptions of assessment trustworthiness. Epistemologically, the knowledge which is generated by the present research is contextualised and socially constructed; it is, furthermore, yielded from interviews with the study participants. Grounded in the participant-centric approach to contribute to the critical discourse on maintaining the trustworthiness of assessment, the research focuses on the EFL instructors who encounter a plethora of educational challenges from a subjective perspective. Methodologically, this research adopted the qualitative method (case study design) to explore certain phenomena using deep investigation of specific cases. Opting for the qualitative methodology is inspired by the research purpose of uncovering the multifaceted dimensions of how to achieve EFL assessment trustworthiness focusing on the challenges and problems encountered by the EFL instructors while assessing their students. The qualitative methodology is suited for investigating the ethical concerns such as biases and contextual challenges while ensuring methodological rigor through careful data collection and contextual analysis.

1. Participants of the research

The population of the present research is the EFL staff members. The study participants were nine EFL instructors who were purposively selected based on their experiences in the field of EFL instruction and were familiar with AI technologies to fulfill the purpose of the research (Patton, 2002). The participants were males and females from diverse nationalities, namely three Egyptians, three Saudis, two Jordanians, and one Sudanese, to gain comprehensive insights from different educational perspectives in the Arab world. The participants were purposively selected considering some criteria: EFL instruction experience, familiarity with AI tools, and willingness to partake in the research using the convenience sampling technique, which was favoured by Etikan et al. (2016) because of its effectiveness in qualitative research.

2. The semi-structured interview

In line with the interpretivism ontology, data were collected from nine EFL instructors using semi-structured interviews to thoroughly explore their experiences, perceptions and insights on the problems encountered, and the solutions. Interview, as a self-contained method (Morgan, 1997), was used as a chief source of data collection to collect high-quality specialised data in a context regarding the undergone construct. The semi-structured interviews were employed due to their flexibility and depth of data collection. The interview process went through a systematic process starting from planning the interview, identifying the objectives, contacting the participants, identifying the allocated time and place, and preparing some prompt questions (Anderson, 1990). The next phase included selecting the participants (purposive sample) and finally, recording the interviewees' responses via notetaking and a digital recorder. The participants' consent was obtained before conducting and recording the interview to align with the conventions of scientific research ethics. The interviewer's role was the moderator, and the semi-structured interviews were individually conducted with the participants. Moreover, the interviews were conducted face-to-face and online.

3. Data analysis

Employing Creswell (2016) model of qualitative data analysis, the following procedures were followed. Firstly, the listening tracks were carefully listened to by both researchers; then, the audio tracks were transcribed verbatim and were checked by both researchers interchangeably word by word, phrase by phrase and sentence by sentence. The process of generating themes passed through three stages, namely open coding, axial coding, and selective coding (main themes).

To ensure the credibility of the interview, the audit trail, an expert of the project (auditor) was formally brought into the research. The audit was the other researcher, as both researchers worked as auditors for each other's work to examine both the process and product of inquiry and determine the trustworthiness of the findings. Through this process of reviewing and documenting the study, the credibility of the qualitative output was assured.

4. Results and Discussion

The current research was developed to address a two-fold purpose, namely exploring the challenges of the unethical employment of AI tools by the EFL majors and identifying some proposed solutions. Keeping this in mind, the following section tackles the first question: "What challenges do EFL instructors encounter that may affect the trustworthiness of assessments?" The qualitative analysis of the semi-structured interviews revealed the following themes addressing the challenges of unethical employment of AI tools used by EFL majors.

Challenges of AI Tools Employment

- *AI Tools and Academic Integrity*

Employing AI tools to generate the EFL learners' work, such as writing essays, translating assignments and answering the reading comprehension questions, negatively affects academic integrity. Not only do such practices reduce the value of the learning experience, but they also damage the traits of honesty and accountability among EFL students when submitting work that is not theirs. The usage of

such tools poses a plethora of thoughts about the authenticity of the work provided. A study respondent stated, *"Depending on the AI tools for performing their tasks, such as translation, writing, and reading, makes them unable to think or do anything."* Another respondent asked, *"How can I identify that the students' production is authentic or not. It is a real dilemma."* A third respondent opined, *"So, the work provided is honest work or not? a perplexing situation."*

- *AI Tools and Students' Engagement*

Using such tools, which provide quick answers and solutions around the clock, might negatively impact the students' engagement. In other words, when EFL students depend on these tools for accomplishing their assignments and tasks, they do not pay attention to their instructors' explanations. Being immersed or at least engaged in the teaching process is the source of deep comprehension and academic growth. Full dependence on such tools might deprive the students of developing abilities such as analysing, synthesizing, and evaluating, which are necessary components of lifelong learning. One of the participants stated, "*The students' motivation and interaction inside the class vanished since the adoption of the AI tools. No one needs to make use of his abilities but a ready-made solution.*" Another respondent stated, "*The students skip their lectures, and AI tools do their assignments easily …*" A third interviewee mentioned "*So, the students are not learning nor developing.*"

- *AI Tools and Students' Independence*

Being independent means holding the responsibility of learning. Furthermore, building confidence in their abilities and possessing self-discipline is a critical quality that must be possessed by 21st century EFL learners. The dependence on such AI tools might result in passive EFL learners who are most of the time waiting for the suggestions and answers of the AI tools instead of functionalising their abilities and improving them. Being able to accurately plan, implement and evaluate the outcomes is not practiced when AI tools are used, as they do all these jobs for the EFL learners in no time. A staff member said, *"I can say that the students are not ready to make use of their minds, but they can wait for the AI tools to provide them with the suggestions, solutions, answers or what they want."* Another respondent remarked, *"Students are not sure about the work they are submitting and have no self-confidence to present their ideas in front of their classmates."* A third instructor mentioned, *"Honestly, I can say that I do not see the majority of students take responsibility for their learning."*

- *AI Tools and the Higher Mental Skills*

Critical thinking and problem-solving are part and parcel of the outcomes of every successful learning process to be used in real-life scenarios. The sub-skills of critical thinking, for instance, information analysis, evidence evaluation and decision making, are not developed when using AI tools to provide a professionally well-developed end-product. On the same line, using AI tools to solve a problem means ignoring a plethora of sub-skills such as identifying the problem, providing solutions, and testing solutions; instead, EFL learners opt for ready-made solutions for the problems. Ignoring such skills means superficiality of the outcomes of the teaching-learning stage and on the job market eventually. An interviewee stated, *"Students use AI for fulfilling every task they want whether simple or complex."* Another respondent confirmed, *"How can those who depend on AI for generating responses be able to think properly or solve educational problems."*

- *AI Tools and Traditional Assessment Practices*

The conventional assessment practices, which mostly depend on the final product, have been the predominant forms of assessment for ages. Such forms of assessment proved to be limited as they only assess outcomes regardless of the process(es) employed. The planning, drafting, and revision are all necessary steps that should be mastered by EFL learners; however, such steps are bypassed by AI tools for the final product and outcomes. One of the participants stated, *"The traditional assessment of writing focuses on the end product, but the process is not assessed at all."* Another respondent confirmed, *"AI tools provide perfect writing tasks that are well written and revised with no mistakes."* Another participant concurred, *"The students do not make drafts or revise their work."*

- *AI Tools and the Students' Creativity*

Creativity and diversity of thought are two integral components of human teaching-learning outcomes. Possessing the ability to generate innovative original ideas based on the human mind's creativity, not the repetitive predicted responses of the AI, which are controlled by the suitability of the prompts given, is an essential attribute of 21st-century EFL learners. The negative dependence on AI can produce monotonous repetitive ideas which cannot be applicable in some contexts; or such ideas are idealized so as not to be suitable in most real contexts. A respondent stated, *"When comparing a lot of work which is produced by ChatGPT, I feel that the ideas are the same but rephrased."* Another participant suggested, *"Sometimes, AI is producing ideas that are funny and not related to the topic at all."* A third respondent mentioned, *"The students sometimes submit content produced by AI tools without modifications or revisions related to different cultures and environments."*

- *AI Tools and the Assessment Metrics and AI Detection Tools*

The worst problem encountered by EFL educators is the inability of the assessment metrics to uncover the sections that are completed using AI tools. Doubting the integrity and fairness of the assessment is a serious problem, as it may hinder the accurate identification of both strengths and weaknesses, thereby negatively impacting assessment practices. The free traditional software detectors are not supported with mechanisms and resources that provide a clear image of the work under investigation. Useless assessment practices do not give feedback to all those who are involved in the teaching-learning process. A respondent mentioned, *"I have tried myself to ensure the*

authenticity of a piece of writing, but AI detectors gave different results from each other." Another participant confirmed, *"I think unpaid detectors are unreliable."* A third interviewee asserted, *"Some AI tools can rephrase and humanise the content to be similar to that of the human".*

- *AI Tools and the Instructors' Training*

Effective employment of AI tools in the teaching and learning process necessitates training and empowering instructors with the necessary knowledge to ensure proper implementation and avoidance of any negative consequences. On the same line, utilizing other platforms, which are based on AI services to detect the content developed by AI tools is another critical skill. Some instructors are not familiar enough with the knowledge and skills needed to appreciate the value and benefits of using such tools appropriately. Comprehensive and continuous professional development is urgently needed to thoroughly understand the mechanisms and benefits of such tools as well as the drawbacks. Enhancing the instructors' competencies in the usage of AI also means that the instructors should not resist the change but take constructive actions. One of the participants concurred, *"AI is developing very fast, and every day there are a lot of new tools that students can use."* Another respondent assured, *"Providing extensive, specialised, and continuous training is very important for instructors to be able to understand the new developments and avoid their negative effects."* A third participant confirmed, *"Traditional training is not a solution at all. Technology is developing every second."*

Proposed Solutions

The following section tackles the second research question "What solutions can address the challenges faced by the EFL instructors for ensuring trustworthy assessments?" The qualitative analysis of the semi-structured interviews revealed the following themes, proposing some practical solutions for overcoming the challenges of AI tools and the unethical employment by the EFL majors.

- *Assessing Higher Order Thinking Skills*

On the other hand, making use of assessment practices that focus on assessing higher mental skills such as problem solving, critical thinking skills, as well as creativity might serve a two-fold purpose. Firstly, AI tools would not be able to imitate the outcomes easily; secondly, such practices encourage deep engagement with the teaching-learning process as well as creativity and authenticity, which means the assessment process reflects the EFL learner's knowledge and skills. The assessment techniques ought to be regularly reviewed to ensure their reliability, validity, and resistance to be tackled by AI tools. Using, for instance, assessment activities that require higher order thinking to be accomplished using authentic language might not be easily processed by AI tools. One of the participants opined, *"If the assessment included questions which target higher mental skills such as problem-solving, content analysis, peer-reviewing, etc…, I think the students will not be able to use AI to complete these tasks."* Another respondent stated, *"I think AI tools cannot do tasks which target critical thinking or problem-solving."* A third interviewee pinpointed, *"The tasks must be of different stages and assess many skills to be difficult for the AI tools to accomplish."*

- *Adopting Active Learning and Assessment Strategies*

Making use of active learning strategies that boost engagement and meaningful interaction in the teaching-learning process might be a feasible solution for overcoming detachment in the educational process. Creating chances for EFL learners to interact and use the language effectively, not only in the teaching-learning process but in the assessment practices, might guarantee collaborative and professional involvement. Such practices enhance experiential learning activities and cultivate concentration on the subject matter and the skills rather than the dependence on the automated responses of the AI tools. An interviewee asserted, *"The assessment tasks must focus on the use of the language not just content production."* Another respondent highlighted, *"Using strategies that encourage collaborative activities and peer work must help in overcoming the excessive use of AI by students." A third participant mentioned,* "*If the assessment focuses on usage, AI tools will not be helpful to students.*"

- *Developing Training Programs for Promoting Assessment Ethics Use*

Implementing workshops and training programs for addressing academic integrity focusing on the unethical employment of AI tools in EFL learning and other subjects; instilling the ethics of integrity and the significance of attributing sources to their founders and avoiding plagiarism; getting EFL students engaged in critical reflection about the ethical consequences of the dishonest use of AI tools encourage them to make informed decisions about using such tools; providing academic modules about AI unethical use might raise the EFL learners' awareness about the accountable and honest usage of AI tools; implementing case studies and discussions about the negative long-run consequences of automation on education are significant contributors to raising EFL students' awareness about the negative consequences of using AI tools unethically. Such courses advocate for ethical AI use and increase EFL students' awareness about the ethical usage of AI tools. One of the respondents mentioned, *"From my experience, when I talked with the students about the unethical use of the AI, many of them were positive about the disadvantageous effect of it, but this is not enough."* Another participant illustrated, *"Some students do not know that AI tools can be harmful to them when they do their work."* Another interviewee confirmed, *"Training programs and open discussions should be held to help students know the harmful effects of AI."* Another participant stated, *"The students must have a full course about AI addressing all its aspects."*

- *Ensuring Ongoing Professional Development*

The rapid development of AI necessitates continuous professional development of educators to familiarise them with the appropriate use of AI tools and ensure the authenticity and integrity of assessment among EFL students. Equipping educators with the necessary resources

to properly integrate AI tools in teaching and assessment practices is of crucial importance. One of the respondents mentioned, *"The promotion courses for the staff members should address a lot of issues about AI."* Another participant stated, *"There must be continuous training workshops and seminars by experts about the effective use of AI technology."* An interviewee highlighted, *"Resources should be available for instructors to help them update their knowledge and skills about AI technology." Another respondent suggested, "AI is developing very fast, and every day there are a lot of new tools which students can use." An interviewee asserted, "Providing extensive, specialized, and continuous training is very important for instructors to be able to understand the new developments and avoid their negative effects."*

- *Establishing Clear Policies and Guidelines*

Identifying clear guidelines and policies is essential to ensure academic integrity among all those involved in the teaching-learning process. Identifying the consequences of misconduct, such as using penalties, and disciplinary actions would ensure transparent and accountable practices among educators. One of the participants stated, *"The university should put clear regulations and penalties for those who use AI unethically."* Another respondent mentioned, "*I think if clear guidelines are given to students about the tasks and the penalties when making something wrong, it will help overcoming this problem.*" An interviewee confirmed, "*Those who violate the rules should be severely punished.*"

- *Updating Traditional Assessment Methods*

Another item of interest is that the outdated traditional assessment practices, which proved untrustworthy in assessing the EFL learners' outcomes, ought to be replaced with other assessment forms. Designing projects, collaborative tasks, presentations, videos, real-life outcomes, etc. might be an effective alternative to the traditional assessment methods that encourage engagement and lessen the dependence on AI tools. Moreover, using assessment techniques that focus on the process rather than the product is significant to ensure the accomplishment of the learning objectives. Likewise, diversifying assessment formats or asking EFL students to accomplish their tasks in two different forms, where each method assessing the other to ensure that the learner is not using AI tools, is another feasible step. On the other hand, implementing peer evaluation as a means of assessment might contribute significantly to updating the dated assessment practices. Not only does the process enhance numerous skills among EFL learners such as critical thinking and problem-solving, but also a sense of responsibility is fostered among EFL learners, apart from assessing the authenticity and the parts which are processed by AI tools. One of the participants stated, *"If we ask the students to submit their work followed by a video about it, it won't make the students depend on the AI."* Another respondent affirmed, *"Using tasks which are not traditional cannot be done by the AI."* A third interview confirmed, *"We must ask the students to include the procedures they followed to reach their answers to ensure that they did not use ChatGPT."*

- *Employing Advanced Detection Tools*

The revolutionary development of AI technology should be aligned with AI detection tools that could detect the parts processed by AI tools with a satisfying degree of reliability. The integration of such detection tools enables the instructors to identify the AI-processed content and the authentic original work devised by the student. The feedback provided by such algorithms might be beneficial for EFL instructors. A respondent confirmed, *"The university should provide the instructors with platforms to help them identify the tasks which are done by ChatGPT."* Another participant asserted, *"Students' assessment must be checked with advanced AI detection tools."* Another interviewee illustrated, *"As there are plagiarism detection tools, there must be strong AI detection tools."*

5. Discussion of the Results

The research findings underscored substantial challenges posed by the EFL instructors, which impact the academic integrity due to the employment of the traditional assessment tools in the age of AI. The generation of essays, translations, and answers to questions raises doubts about the extent to which the assessment tools are effective in providing a valid overview of the learning outcomes; such findings are in line with that of Holmes et al. (2021) and Zawacki-Richter et al. (2019). More significantly, the excessive use of technology would stimulate students to opt for ready-made options that may hinder the development of many linguistic and thinking skills and lead to passive learning. Focusing on the learning process rather than the final product to assess the higher order thinking skills necessitates creative teaching strategies and valid assessment tools (Anderson and Krathwohl 2001; Wilson, 2016). Furthermore, transparency and accountability, two main significant components of every successful educational practice, might be negatively impacted by the unethical employment of AI tools, which aligns with what was attained by Williamson (2017).

The unethical adoption of AI tools mostly leads to passive learning where students might depend on them to generate answers for all the assignments and projects which contradicts the constructive learning theory (George, 2023) which calls for a meaningful engagement in the teaching-learning process as an essential component for effective knowledge construction and decreases the dependence on the AI tools for accomplishing assignments. AI tools also bypass the higher order thinking skills, e.g. critical thinking, analysis, synthesis, and evaluation as indicated by Bloom's Taxonomy (Krathwohl, 2002). The findings of the study are in line with the validity theory, which was proposed by Messick (1989) and confirmed the significance of designing assessment tools that are fair, reliable and in line with the contemporary challenges. The findings also highlighted that traditional assessment methods are often violated by the work generated by the unethical use of AI tools.

Based on the findings of the current study, there is an urgent need to consider the traditional assessment methods currently used in the educational practices in the EFL context. Thus, authentic assessment as a reforming practice should be implemented to assess higher-order thinking skills such as critical thinking and creativity (Al-Otaibi, 2016). Such a conclusion is aligned with what was indicated by Ghimire and Edwards (2024), who confirmed that students can use AI tools to accomplish programming tasks, and Rossell ó-Geli (2022), who underscored a dire need for assessing AI interference in teaching and assessment. Another urgent issue is the development of a moral code to enhance the EFL learners' awareness regarding the pros and cons of using AI tools. Such code may contribute to increasing academic integrity and decreasing the unethical usage of AI tools to accomplish academic tasks, not to mention alleviating "AI guilt" reported by students when using such tools (Chan, 2024). Furthermore, Han et al. (2023) and Xie et al. (2022) also underscored other psychological consequences of overreliance on AI tools such as learners' autonomy reduction and anxiety increase.

More importantly, employing technological means, such as advanced content detectors, is another relatable item. Al-Jarf (2024) pinpointed that educators reject the tasks accomplished by AI tools, which necessitates the dependence on reliable tools to detect AI use. Another relevant point is that the professional development of staff members ought to comprise mastering such tools and identifying their implications in the teaching-learning process (Zhao, 2024). The adoption of active learning strategies and techniques might contribute to the achievement of learning outcomes and counteracting the negative consequences of using AI tools in accomplishing academic tasks. Authentic interactive classroom activities mean real-world application, collaboration, and engagement, which cannot be conducted by AI tools as quoted by Tosuncuoglu (2018). When linking the teaching practices with increasing the students' responsibility for their learning, the teaching-learning outcomes might be maximised to the optimal level (Moya, 2023; Smolansky, 2023).

6. Research Conclusions

To conclude, while AI tools proved to be useful gadgets in the teaching-learning process, they pose a plethora of severe challenges to academic integrity by generating responses threatening the core values of academic work. AI negatively affects the EFL learners' engagement and impairs critical thinking and creativity among them, leading to passivity, monotony and weak learning outcomes when used unethically. Furthermore, the traditional means of assessment that focus on the product are another challenge encountered by EFL instructors as such traditional methods are easily compromised by AI tools necessitating launching assessment reform procedures for utilizing authentic means that target assessing higher mental skills using innovative techniques such as projects, peer-reviewing, presentations, portfolios, and procedures of completing tasks. More importantly, although the dependence on AI-processed content detection tools might be a reliable means for ensuring originality, continuous professional development as well as raising the EFL students' awareness might be feasible solutions for such challenges posed by AI tools and techniques.

7. Suggestions for Further Research

Building on the findings attained by the present research, various avenues are cited to address the challenges of employing AI in EFL assessment.

- Exploring the resistance of authentic assessment methods, such as projects, oral presentations, and written tests, to AI processing.
- Exploring the ethical implications related to the adoption of AI tools in EFL instruction and assessment.
- Investigating the impact of developing policies and regulations for ethical implementation of AI tools.
- Exploring the impact of AI-augmented feedback on the development of the four language skills.

Acknowledgement

The authors extend their appreciation to Prince Sattam bin Abdulaziz University for funding this research work through the project number (PSAU/2024/02/31331).

The authors would like to express their gratitude to the nine EFL instructors who participated in this study for their valuable input and willingness to share their experiences.

Authors' contributions

Dr. Shaalan handled data collection, analysis, and manuscript drafting. Dr. Ahmad led the study design, methodology, and literature review. Both authors collaborated on the discussion and revisions and approved the final manuscript.

Funding

This work was supported by Prince Sattam bin Abdulaziz University [project number PSAU/2024/02/31331].

Competing interests

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Informed consent

Obtained.

Ethics approval

The Publication Ethics Committee of the Sciedu Press.

The journal's policies adhere to the Core Practices established by the Committee on Publication Ethics (COPE).

Provenance and peer review

Not commissioned, externally double-blind peer-reviewed.

Data availability statement

The data that support the findings of this study are available on request from the corresponding author. The data is not publicly available due to privacy and ethical restrictions.

Data sharing statement

No additional data is available.

Open access

This is an open-access article distributed under the terms and conditions of the Creative Commons Attribution license (http://creativecommons.org/licenses/by/4.0/).

Copyrights

Copyright for this article is retained by the authors, with first publication rights granted to the journal.

References

Abd Elgalil, H. M., Abd El -Hakam, F. E.-Z., Farrag, I. M., Abdelmohsen, S. R., & Elkolaly, H. (2022). Undergraduate students' perceptions of online assessment during COVID -19 pandemic at faculty of medicine for girls, Al-Azhar University, Cairo, Egypt. *Innovations in Education and Teaching International, 60*(2), 185–195. https://doi.org/10.1080/14703297.2022.2037450

Abd El-Magid, A. D. S. (2024). Applications of artificial intelligence in education (developments and future visions): A review. *Al-Azhar Journal of Education, 43*(202), 553-617. https://doi.org/10.21608/jsrep.2024.361882

Abdulrazzaq, D. M. & Abdellatif, M. S. (2023). Online assessment or offline Assessment, which one is more aggressive? The impacts on willingness to communicate, test taking anxiety, and language achievement. *Computer-Assisted Language Learning Electronic Journal, 24*(2), 69-85.

Ahmad, A. S. K., El-Helaly, A. S., & Al-Tanany, A. A. (2021). The effectiveness of 4MAT model via Google Classroom in developing argumentative writing skills among EFL majors. *Al-Azhar Journal of Education*, *40*(192), 227-259. https://doi.org/10.21608/jsrep.2021.314156

Al Ghamdi, F. K., & Al-Ghamdi, S. M. M. (2021). The effectiveness of a training program based on constructive learning model in developing authentic assessment among science female teachers. *Al-Azhar Journal of Education*, *40*(189), 326-382. https://doi.org/10.21608/jsrep.2021.171755

Al-Jarf, R. (2024). Students' Assignments and Research Papers Generated by AI: Arab Instructors' Views. *Journal of Computer Science and Technology Studies, 6*(2), 92-98. https://doi.org/10.32996/jcsts.2024.6.2.11

Al-Otaibi, S. S. A. (2016). A proposed concept for the use of computer teachers of electronic portfolios (E-Portfolio) in evaluating the performance of first-year secondary students in light of alternative assessment. *Al-Azhar Journal of Education*, *35*(170), 753–822. https://doi.org/10.21608/jsrep.2016.33857

Amua-Sekyi, E. T. (2016). Assessment, Student Learning and Classroom Practice: A Review. *Journal of Education and Practice, 7*(21), 1-6. Retrieved from https://files.eric.ed.gov/fulltext/EJ1109385.pdf

Anderson, G., (1990). *Fundamentals of educational research.* The Falmer Press.

Anderson, L. W., & Krathwohl, D. R. (2001). *A taxonomy for learning, teaching, and assessing: A revision of Bloom's taxonomy of educational objectives: complete edition*. Addison Wesley Longman, Inc.

Atlas, S. (2023). *ChatGPT for higher education and professional development: A guide to conversational AI.* University of Rhode Island

Chan, C. K. Y. (2024). *Exploring the Factors of" AI Guilt" Among Students--Are You Guilty of Using AI in Your Homework?* https://doi.org/10.48550/arXiv.2407.10777

Chandler, D. (1995). *Technological or Media Determinism*. Retrieved from https://shaunlebron.github.io/chandler-1995.pdf

Chen, L., Chen, P., & Lin, Z. (2020). Artificial Intelligence in Education: A Review. *IEEE Access, 8*(1), 75264-75278. https://doi.org/10.1109/access.2020.2988510

Cheng, L., Rogers, T., & Hu, H. (2004). ESL/EFL instructors' classroom assessment practices: Purposes, methods, and procedures. *Language Testing, 21*(3), 360-389. https://doi.org/10.1191/0265532204lt288oa

Creswell, J. (2016). *Research design: Qualitative, quantitative, mixed methods approaches.* University Of Nebraska-Lincoln.

Driscoll, A., & Wood, S. (2023). *Developing Outcomes-Based Assessment for Learner-Centered Education.* Routledge. https://doi.org/10.4324/9781003444176

Etikan, I., Musa, S. A., & Alkassim, R. S. (2016). Comparison of convenience sampling and purposive sampling. *American journal of theoretical and applied statistics*, *5*(1), 1-4. https://doi.org/10.11648/j.ajtas.20160501.11

George, A. S. (2023). Preparing students for an AI-driven world: Rethinking curriculum and pedagogy in the age of artificial intelligence. *Partners Universal Innovative Research Publication*, *1*(2), 112-136. https://doi. org/10.5281/zenodo.10245675

Ghimire, A., & Edwards, J. (2024, July). Coding with ai: How are tools like ChatGPT being used by students in foundational programming courses. In *International Conference on Artificial Intelligence in Education* (pp. 259-267). Cham: Springer Nature Switzerland. https://doi.org/10.1007/978-3-031-64299-9_20

Gultom, E. (2016). Assessment and evaluation in EFL teaching and learning. *Proceedings of ISELT FBS Universitas Negeri Padang*, *4*(1), 190-198. Retrieved from https://ejournal.unp.ac.id/index.php/selt/article/view/6928/5462

Gusev, M., & Armenski, G. (2014). E-assessment systems and online learning with adaptive testing. *E-Learning Paradigms and Applications: Agent-based Approach*, 229-249. https://doi.org/10.1007/978-3-642-41965-2_8

Habib, S., Vogel, T., Anli, X., & Thorne, E. (2024). How does generative artificial intelligence impact student creativity? *Journal of Creativity*, *34*(1), 100072. https://doi.org/10.1016/j.yjoc.2023.100072

Han, B., Nawaz, S., Buchanan, G., & McKay, D. (2023, June). Ethical and pedagogical impacts of AI in education. In *International Conference on Artificial Intelligence in Education* (pp. 667-673). Cham: Springer Nature Switzerland. https://doi.org/10.1007/978-3-031-36272-9_54

Holmes, W. (2019). *Artificial intelligence in education: Promises and implications for teaching and learning.* Center for Curriculum Redesign.

Hristov, K. (2016). Artificial intelligence and the copyright dilemma. *Idea*, *57*, 431. https://doi.org/10.2139/ssrn.3490458

Ibrahim, H., Asim, R., Zaffar, F., Rahwan, T., & Zaki, Y. (2023). Rethinking homework in the age of artificial intelligence. *IEEE Intelligent Systems*, *38*(2), 24-27. https://doi.org/10.1109/mis.2023.3255599

Javed, F. (2024). The Evolution of Artificial Intelligence in Teaching and Learning of English Language in Higher Education: Challenges, Risks, and Ethical Considerations. In *The Evolution of Artificial Intelligence in Higher Education: Challenges, Risks, and Ethical Considerations* (pp. 249-276). Emerald Publishing Limited. https://doi.org/10.1108/978-1-83549-486-820241015

Jonassen, D. H. (1999). Constructivist learning environments on the web: Engaging students in meaningful learning. *Educational Technology*, *39*(3), 35-45. Retrieved from https://tinyurl.com/2vxmd726

Krathwohl, D. R. (2002). A revision of Bloom's taxonomy: An overview. *Theory into Practice*, *41*(4), 212-218. https://doi.org/10.1207/s15430421tip4104_2

Krzykowski, L., & Kinser, K. (2014). Transparency in student learning assessment: Can accreditation standards make a difference? *Change: The magazine of higher learning*, *46*(3), 67-73. https://doi.org/10.1080/00091383.2014.905428

Kushari, B., & Septiadi, L. (2022). A learning outcome assessment information system to facilitate Outcome-Based Education (OBE) implementation. *Jurnal Pendidikan Teknologi dan Kejuruan*, *28*(2), 238-250. https://doi.org/10.21831/jptk.v28i2.42339

Lindsay, E. D., Johri, A., & Bjerva, J. (2023). A Framework for Responsible Development of Automated Student Feedback with Generative AI. https://doi.org/10.48550/arXiv.2308.15334

Marzano, R. J. (2010). Formative versus summative assessments as measures of student learning. In *Handbook of data-based decision making in education*. Routledge.

Meek, T., Barham, H., Beltaif, N., Kaadoor, A., & Akhter, T. (2016, September). Managing the ethical and risk implications of rapid advances in artificial intelligence: A literature review. In *2016 Portland International Conference on Management of Engineering and Technology (PICMET)* (pp. 682-693). IEEE. https://doi.org/10.1109/picmet.2016.7806752

Menard, L. (2021). Personalizing Learning Targets with Technology-Based Assessment. In *Handbook of Research on Critical Issues in Special Education for School Rehabilitation Practices* (pp. 278-303). IGI Global. https://doi.org/10.4018/978-1-7998-7630-4.ch015

Messick, S. (1989). Validity. In *Educational measurement* (3rd ed., pp. 13-103). American Council on Education.

Morgan, D. L. (1997). *The focus group guidebook*. Sage Publications. https://doi.org/10.4135/9781483328164

Moya, B., & Eaton, S. E. (2023). Examining Recommendations for Artificial Intelligence Use with Integrity from a Scholarship of Teaching and Learning Lens. *RELIEVE-Revista Electrónica de Investigación y Evaluación Educativa*, *29*(2), 1-21. https://doi.org/10.30827/relieve.v29i2.29295

Musa, H. I. A. (2023). Dialogic Inquiry: A Tool for Developing Oral Communicative Competence and Increasing the Willingness to Communicate among EFL Learners. *World Journal of English Language*, *13*(2), 278-278. https://doi.org/10.5430/wjel.v13n2p278

Patton, M. Q. (2002). *Qualitative evaluation and research methods* (3rd ed.). Sage.

Playfoot, D., Quigley, M., & Thomas, A. G. (2024). Hey ChatGPT, give me a title for a paper about degree apathy and student use of AI for

assignment writing. *The Internet and Higher Education, 62*(100950), 1-20. https://doi.org/10.1016/j.iheduc.2024.100950

Romano, L. E. (2024). Assessment for Equity: Exploring How Secondary Educators Utilize Classroom Management and Assessment Practices to Sustain Student Identities. *Assessment for Effective Intervention, 49*(2), 75-85. https://doi.org/10.1177/15345084231178788

Rossell ó-Geli, J. (2022, April). Impact of AI on Student's Research and Writing Projects. In *International Conference on Computational Intelligence in Pattern Recognition* (pp. 705-713). Singapore: Springer Nature Singapore. https://doi.org/10.1007/978-981-99-3734-9_57

Shaalan, I. E. N. A. W., & Ahmad, A. S. K. (2024). Linguistic competence among Egyptian vs. Saudi EFL majors in light of utilizing artificial intelligence technology: A predictive study. *International Journal of Computer-Assisted Language Learning and Teaching (IJCALLT), 14*(1), 1-19. https://doi.org/10.4018/ijcallt.361771

Smolansky, A., Cram, A., Raduescu, C., Zeivots, S., Huber, E., & Kizilcec, R. F. (2023, July). Educator and student perspectives on the impact of generative AI on assessments in higher education. In *Proceedings of the tenth ACM conference on Learning@ Scale* (pp. 378-382). https://doi.org/10.1145/3573051.3596191

Suskie, L. (2018). *Assessing student learning: A commonsense guide*. John Wiley & Sons.

Tosuncuoglu, I. (2018). Importance of Assessment in ELT. *Journal of education and training studies, 6*(9), 163-167. https://doi.org/10.11114/jets.v6i9.3443

Wei, W. (2015). Using summative and formative assessments to evaluate EFL teachers' teaching performance. *Assessment & Evaluation in Higher Education, 40*(4), 611-623. https://doi.org/10.1080/02602938.2014.939609

Williamson, B. (2017). *Big data in education: The digital future of learning, policy and practice.* SAGE Publications. https://doi.org/10.4135/9781529714920

Wilson, L. O. (2016). Anderson and Krathwohl Bloom's taxonomy revised understanding the new version of Bloom's taxonomy. *The Second Principle, 1*(1), 1-8.

Xie, C., Ruan, M., Lin, P., Wang, Z., Lai, T., Xie, Y., ... Lu, H. (2022). Influence of artificial intelligence in education on adolescents' social adaptability: A machine learning study. *International journal of environmental research and public health, 19*(13), 7890. https://doi.org/10.3390/ijerph19137890

Yeo, M. A. (2023). Academic integrity in the age of Artificial Intelligence (AI) authoring apps. *TESOL Journal, 14*(3), e716. https://doi.org/10.1002/tesj.716

Yu, H., Jeon, E., Oh, S., Son, D., Lee, S., & Lim, K. S. (2022, October). Design of Interactive Reading Comprehension Competence Assessment System using AI. In *2022 13th International Conference on Information and Communication Technology Convergence (ICTC)* (pp. 2062-2064). IEEE. https://doi.org/10.1109/ictc55196.2022.9952613

Zawacki-Richter, O., Mar ń, 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. https://doi.org/10.1186/s41239-019-0171-0

Zhao, M. (2024). A study on the measurement and standardized assessment model of student learning outcomes in vocational institutions. *Applied Mathematics and Nonlinear Sciences, 9*(1). 1-117. https://doi.org/10.2478/amns.2023.2.00400

Zohny, H., McMillan, J., & King, M. (2023). Ethics of generative AI. *Journal of medical ethics*, *49*(2), 79-80. https://doi.org/10.1136/jme-2023-108909