Artificial intelligence-based assessment in ELT exam creation: a case study of Van Lang University lecturers

Duy Nguyen Luc Ha and Anh Tu Nguyen
Institute of Postgraduate Education, Van Lang University, Ho Chi Minh City, Vietnam

Abstract

Purpose – Focusing on the growth of artificial intelligence (AI) in education, this research reveals that AI can create and improve English language assessments for learners in order to optimize and enhance test questions as a bilateral tool in the traditional way for English Language Teaching (ELT) is possible.

Design/methodology/approach – The research adopted a qualitative methodology by conducting semi-structured interviews with a varied range of language institutes' lecturers, revealing new beneficial effects of AI on test time, content and human variables.

Findings – Several interviewees agreed that AI should be used in ELT exam creations because of its overt advantages in making test items automatically, adaptive testing, enhanced feedback mechanisms and quality assurance and innovative formats. Simultaneously, some disadvantages are recorded, including complexity and nuance of language, technical limitations, ethical and bias concerns and human oversight and validation.

Research limitations/implications – The study was also limited by the time frame of the research, which may not have fully captured the complex dynamics between the different actors, such as using AI in preparing questions for reading tasks such as automatic creation of pre-reading questions as well as possible answers.

Originality/value – For future studies, as AI-generated material is becoming more ubiquitous, from music to artwork, it presents crucial legal problems regarding who owns the rights to the work or construct ELT exams. It has also become the next problem that the writers should concentrate on.

Keywords Van Lang University, English language teaching (ELT), Artificial intelligence (AI), English language assessments, Test creation

Paper type Case study

Introduction

Assessment plays a pivotal role in the educational training process, offering essential insights into the effectiveness of language instruction by identifying challenging areas and gauging learner progress in language courses (Pearson and Murphy-Judy, 2020; Voss, 2018). It also enables lecturers to assess their students' language abilities, pinpoint strengths, and weaknesses, and propose targeted improvements (Brown and Abeywickrama, 2010; Purpura, 2016). Commonly used tools such as multiple-choice questions, reading comprehension tasks, and oral language assessments (Hughes and Hughes, 2020) require significant time and effort to develop, as they must align with learners' current needs and desired outcomes. This time-intensive process is crucial to ensuring precise and effective language proficiency evaluations at the university level.

Despite the recognized value of language assessment, lecturers often face the burden of developing language test items that align with institutional or national curriculum standards. The time and energy invested in these efforts are essential for fostering future

student success and ensuring quality education. However, recent technological advances, particularly in artificial intelligence (AI), have transformed the educational landscape by streamlining time-consuming tasks (Çakmak, 2019; Kukulska-Hulme and Morgana, 2021; Selwyn *et al.*, 2021). In language education, technology has significantly enhanced teaching and learning practices, offering tools to create assessments automatically and adaptively (Nazaretsky *et al.*, 2022). Although AI presents substantial opportunities for improving assessment efficiency and adaptability (García-Peñalvo *et al.*, 2020; Gardner *et al.*, 2021), research exploring the potential risks, hidden errors, and limitations of AI-generated language tests is limited (Van Moere and Downey, 2016). This study seeks to address this gap by examining the unique challenges and opportunities faced by language institute lecturers when using AI to generate test questions and banks for English Language Teaching (ELT) exams. A key contribution of this research is the identification of AI's dual role – not only as a tool for automating test creation but also as a partner in enhancing test content, reducing human biases, and personalizing assessments to meet diverse learner needs. The study provides novel empirical insights into both the practical benefits and the potential pitfalls of AI in ELT assessment, offering a balanced perspective on its integration into the educational process.

Literature review
Language testing and assessment (LTA)
Language Testing and Assessment (LTA) is an integral part of educational systems, aimed at enhancing the quality and efficacy of language teaching and learning (Heaton, 1988). LTA encompasses designing and evaluating assessment procedures to monitor, grade, and assess learners' knowledge and progress (Heaton, 1988; Vogt and Tsagari, 2014; Lan and Fan, 2019). Lan and Fan (2019) trace the origins of LTA to its role in identifying and bridging knowledge gaps among students, thus guiding them toward language acquisition. In contemporary education, it is essential for lecturers to align their assessments not only with theoretical standards but also with practical criteria to ensure that they accurately reflect learners' capabilities and areas for improvement (Stiggins, 2006; Tsagari and Vogt, 2017). Despite the significance of LTA, challenges persist in aligning assessments with teaching due to unpredictable outcomes and various influencing factors, including resources, materials, and lecturer expertise (Bachman and Adrian, 2022; Lam, 2015). Recent empirical studies reveal gaps in the adequacy of test questions and their alignment with students' language levels (Lam, 2015; Tsagari and Vogt, 2023). To address these challenges, LTA needs to be updated and innovated by integrating new factors and methodologies that reflect real-world complexities rather than relying solely on theoretical models (Bachman and Adrian, 2022). This evolution will lead to more timely and accurate assessments that address the gaps between learners' abilities and instructional intent.

Artificial intelligence (AI) in test creation
Artificial Intelligence (AI), as defined by McCarthy *et al.* (2006), involves machines created through scientific and engineering processes that can deliver valuable results across various fields, including education (Helm *et al.*, 2020). AI systems are designed to simulate human thought and action within a virtual scope, enhancing various educational applications (Eaton *et al.*, 2021; McCarthy *et al.*, 2006). In education, AI has introduced innovative tools that significantly impact language teaching and assessment. For example, AI-powered tools like Grammarly and QuillBot assist in grammar checking and paraphrasing, respectively (Godwin-Jones, 2022; Zhang and Zou, 2022). Moreover, AI facilitates automatic grading and feedback on student work, streamlining the assessment process (Gardner *et al.*, 2021; Borade and Netak, 2021; Yu *et al.*, 2022). AI's ability to generate a large volume of diverse test questions quickly improves the efficiency and effectiveness of assessment creation (Hinkelman, 2018).

Research indicates that AI algorithms can produce varied question formats, such as multiple-choice questions, that cater to different testing needs (Killawala *et al.*, 2018). Yunjiu *et al.* (2022) highlight that AI can create diverse test formats, which contrasts with human test creators' tendency to produce repetitive questions based on memory. This capability underscores the potential of combining AI's generalization with human expertise to enhance test content and format diversity (Chen *et al.*, 2023).

AI in test creation of Van Lang university

At Van Lang University, the use of AI in test creation has become increasingly prevalent, offering numerous benefits in terms of speed and quality. AI systems at the university generate customized test items based on extensive educational resources, syllabi, and learning goals, thereby streamlining the test development process (Lee and Kim, 2020; Binh *et al.*, 2024). AI-driven platforms enable lecturers to create diverse and high-quality testing items, aligning with curricular objectives and saving valuable time (Johnson *et al.*, 2019). Furthermore, AI can tailor test questions based on student performance data, providing personalized assessments that enhance learning outcomes (Alshammari and Al-Enezi, 2024), at the same time, noting that Algorithmic biases and data privacy issues must be addressed to ensure equitable and secure assessments. Despite these advantages, several challenges accompany the use of AI in test creation. Concerns about the fairness and validity of AI-generated test questions, lecturers may need training to effectively use AI-driven tools (Rodriguez *et al.*, 2022). The potential for adaptive assessments, which adjust difficulty based on student performance, represents a promising area for future development (Wang *et al.*, 2021). The existing literature on AI in language assessment often highlights the technology's potential to enhance the quality and efficiency of test creation. However, there remains a significant gap in empirical research concerning its practical applications and associated challenges. Current studies have not thoroughly explored how AI can address issues such as human-related biases, improve the richness of test content, and accelerate the test creation process.

Research questions

In line with this aim, the following research questions were proposed:

(1) How might artificial intelligence be used to develop English Language Teaching (ELT) test questions in terms of quality, content richness, and efficiency?

(2) What challenges do lecturers encounter when using artificial intelligence to generate ELT test questions?

Methods

Design of the study

The present work is a case study, a kind of qualitative research, in the form of semi-structured interviews to explore the answers in open-ended to explore valuable and unexpected data from the participants. The purpose is to investigate the lecturers' experiences, opinions, and attitudes about using AI-based evaluation in the context of creating exams. Through an in-depth examination of an individual's experience, emotions, and perceptions, complicated social phenomena were to be explored, analyzed, and explained. Because case studies gather and analyze data from many sources, they enable researchers to comprehend a particular individual, group, or event. The careful examination of the data gathered and the researcher's participation in group projects throughout the study was essential to the design.

Research participants

The study included seven lecturer groups working at the Institute of Language, Van Lang University, which are located at Campus 3 (*Known as Main Campus*) and are ranked their

English courses from Level 1 to Level 7; this data is presented in Table 1 below as well. A purposive sample procedure was utilized in selecting participants so that the objectives of the study were met. For each group, three full-time lecturers were assigned to teach courses number 27, 28, and 29 in the fall semester of the academic year 2023–2024.

To conduct this, strict inclusion criteria were applied to enhance the validity and reliability of the study outcome. It was an obligation to have qualified lecturers with a master's degree or more, over 5 years of teaching experience, and be active members of the English Language Teaching (ELT) examination board. These criteria were designed on the assumption that well-experienced and trained teachers would provide a better and wider view of modern teaching methods and assessment methods according to the literature reviewed (Smith and Fletcher, 2020; Johnson and Johnson, 2020).

The purposive sampling method was considered suitable for this study, as it facilitated the selection of participants with specialized knowledge pertinent to the research objective (Palinkas *et al.*, 2015). Such a sampling strategy was used because, in this study, its only function is to pick individuals who have expertise that is relevant to the focus of the research (Palinkas *et al.*, 2015).

This was crucial for examining AI-assisted learning and assessment techniques, as well as assessing conventional and alternative testing methods within the educational framework. To safeguard participant anonymity, all personal data gathered during surveys and interviews were anonymized in accordance with ethical research standards (Bryman, 2016). Thus, this was very important when addressing AI enabled learning and assessment strategies, as well as testing conventional and alternative approaches for evaluating and administering.

During the research process, the researcher assures strict adherence to the study design requirements, which include the following:

Connectivity: The researcher constructs the survey questions and questionnaires in alignment with the study goals. The questionnaire primarily addresses the variables of ELT's assessment process among participants.

Generalization: The study's results should apply to a larger population because of the research's extensive breadth. To sum up, this study's results include broader demographic characteristics.

Feasibility: Every investigation is constrained by restricted resources. If the research design is beyond the available resources and data accessibility, it is impractical to conduct.

Data collection and analysis
The data-collecting procedure for this research included many essential stages to obtain thorough insights into the use of artificial intelligence (AI) in the design of English Language Teaching (ELT) exams at Van Lang University. The interview questions asked are listed in Appendix 2, depending on the interview, the order may be different. Most interviews were recorded on tape. Then, it was transcribed and checked by the authors. After that, a summary of

Table 1. Seven group of lecturers at the institute of language at Van Lang university

Respondents	Title	Teaching campus	TEFL's level	Interview time
A	Full-time Lecturer	Campus 3	Level 1	45 min
B	Full-time Lecturer	Campus 3	Level 2	55 min
C	Full-time Lecturer	Campus 3	Level 3	60 min
D	Full-time Lecturer	Campus 3	Level 4	70 min
E	Full-time Lecturer	Campus 3	Level 5	80 min
F	Full-time Lecturer	Campus 3	Level 6	75 min
G	Full-time Lecturer	Campus 3	Level 7	75 min

Source(s): Collected by authors

each interview with each participant organized by themes and sub-themes was drafted. This study adapted a structure proposed by Gioia *et al.* (2013) and Jaskiewicz *et al.* (2015) to synthesize interview transcripts and generate main themes and sub-themes. Summaries of the individual interviews with informed consent forms were then sent to each participant for verification (Appendix 1).

Data collection

Semi-structured interviews were conducted with lecturers from the Institute of Language at Van Lang University. Each interview, lasting between 45 and 80 min, was recorded, transcribed, and then summarized to capture key themes and insights. The analysis process adhered to the framework proposed by Gioia *et al.* (2013) and Jaskiewicz *et al.* (2015), which involves a rigorous coding procedure to identify themes and sub-themes within the interview data. This approach facilitated a detailed exploration of lecturers' experiences and perceptions regarding AI-assisted exam development.

To ensure consistency and reliability, interview guidelines were developed with open-ended questions designed to elicit comprehensive responses on AI's role in test creation, including its impact on test item generation, grading, and feedback provision. After transcribing the interviews, the data was coded using an iterative approach: familiarization with the data, initial coding, theme identification, and synthesis (Braun and Clarke, 2006; Saldana, 2015). This process involved generating themes and sub-themes to reflect the lecturers' viewpoints on the use of AI in English Language Teaching (ELT) assessments.

The thematic analysis was complemented by data triangulation, where examples of both traditional and AI-generated test materials were analyzed to provide a broader understanding of AI's impact (Noble and Heale, 2019). Summaries of individual interviews were sent to participants for verification, ensuring the accuracy of the data representation and enhancing the credibility of the findings (Lincoln and Guba, 1988). This comprehensive approach not only illuminated recurring patterns and trends in the lecturers' feedback but also provided actionable insights into the integration of AI in language assessment practices.

Data gathering and analysis

The data analysis for this study employed a rigorous qualitative approach, focusing on thematic analysis to uncover insights from the semi-structured interviews. The analysis adapted a structured process as outlined by Braun and Clarke (2006), which involved several key stages:

The initial step involved thorough reading and re-reading of the interview transcripts to become intimately acquainted with the content. This phase allowed the researchers to immerse themselves in the data and gain an initial understanding of the prevalent issues and patterns.

(1) *Initial coding:* Next, the data were systematically coded. This involved identifying and labeling segments of the text that were relevant to the research questions. Codes were applied to pieces of data that represented significant concepts or ideas related to AI in exam creation. Focusing on specific items such as "AI efficiency," "bias reduction," and "content richness" which were used to tag relevant excerpts from the transcripts.

(2) *Themes searching:* After coding, the researchers aggregated similar codes to form potential themes. This step involved collating all the data relevant to each code and examining how they could be grouped into broader themes. Meaning that, the codes related to "AI efficiency" and "time-saving" were combined into a theme reflecting the efficiency benefits of AI.

(3) *Themes reviewing:* The identified themes were then reviewed and refined. This involved checking whether the themes accurately represented the data and whether they provided a coherent narrative. The researchers iteratively adjusted the themes and subthemes to ensure they captured the essence of the data.

(4) *Themes identification:* Each theme was clearly defined and named to reflect its content and relevance to the research questions. Subthemes were also identified to provide a more nuanced understanding of the data. For instance, within the theme of "AI Efficiency," subthemes such as "automated grading" and "test generation speed" were established to detail specific aspects of AI's impact.

(5) *Interpretation and description:* Finally, the findings were synthesized into a comprehensive report. The themes and subthemes were discussed in relation to the research objectives, and illustrative quotes from the interviews were used to support the analysis. This approach ensured that the interpretation of the data was grounded in the participants' experiences and perceptions.

By following these steps, the study ensured a systematic and transparent approach to analyzing the interview data, which provided rich insights into the role of AI in language assessment and addressed the research questions effectively.

Findings

After the interviews, most interviewees agreed that Artificial intelligence should be used in ELT exam creations because of its overt advantages.

Some of the key features that the authors have listed in Table 2 above clarify the creative characteristics of AI in language testing illustrate the crucial benefits that AI provides are much higher than its limitations. Particularly, the data has elucidated the two sides of AI-based test creation:

RQ1. How might artificial intelligence be used to develop ELT test questions in terms of quality, content richness, and efficiency?

The results indicate several ways that AI might be useful for the education sector's English Language Teaching (ELT) exam question preparation process:

Making test items automatically
The process of creating test questions for English Language Teaching (ELT) examinations can be automated with artificial intelligence (AI).

Groups A, C, and E said that AI can be used to generate testing items that can properly cut down the time spent on exam preparation, emphasizing that Testing items can be taken and prepared in just a few hours.

It is clarified by Group E that the AI tool creates multi-choice questions almost instantaneously, allowing the lecturers can have more time to focus on refining the curriculum and engaging with students.

Table 2. The key benefits provided by lecturers

Item	Category	Percentage of lecturers	Description
Automated test item creation	High Efficiency	30%	"AI tools create multiple-choice questions almost instantaneously"
Adaptive testing	Increased Motivation	25%	"Adaptive tests adjust difficulty based on learner performance"
Enhanced feedback mechanisms	Detailed Feedback	20%	AI provides comprehensive feedback, explaining both correct and incorrect answers
Quality assurance and innovative formats	Improved Test Quality	25%	"AI detects inconsistencies and biases, improving test validity"

Source(s): Collected by authors

In supporting this, Johnson and Lee (2022) give opinions that thanks to Natural Language Processing (NLP) algorithms, AI is able to evaluate vast corpora of text in order to recognize pertinent linguistic patterns and create a wide variety of test questions that are contextually acceptable across a range of language skill levels.

Adaptive testing

The adaptive testing systems driven by AI can modify test materials, serving individual in practicing skills to upgrade their learning process with appropriate levels.

> Lecturers in groups B and D said that Their students seem more motivated during adaptive tests because they don't face a string of discouragingly hard questions. It boosts their confidence to see the challenging questions at an achievable level. This means the questions now can be solved by the student's current knowledge in depth, so they have encouraged motivation to settle the tasks given.

To create a tailored testing experience that effectively tests learners' language abilities and delivers targeted feedback, artificial intelligence may continually analyze learner replies and modify the difficulty level of questions (Garcia *et al.*, 2023). This allows AI to provide personalized testing questions on student capacity.

Enhanced feedback mechanisms

Assessment systems that are based on artificial intelligence can provide learners with rapid and thorough feedback, which assists them in understanding their areas of strength and improvement.

Group lecturers A, C, and G expressed clearly that "the feedback and correction sometimes provided by the AI is incredibly detailed, highlighting not just the correct answers but also explaining why certain responses are incorrect".

> Group A also highlights that "Students appreciate the constructive feedback that goes beyond just marking answers as right or wrong. It helps them understand their mistakes and learn from them."

Therefore, AI is able to give insights into learners' language skill levels, grammatical faults, and vocabulary use via automated grading and analysis of replies (Chen and Smith, 2021). This enables more effective learning tactics to be implemented.

Quality assurance and formats innovative

By identifying inconsistencies, biases, or ambiguities in item development, artificial intelligence has the potential to contribute to implementing quality assurance measures for English Language Testing (ELT) questions.

> According to Group B, and D's statement, "Quality assurance isn't a one-time process. The AI continually analyzes test results and student performance to identify areas for improvement, making our exams better with each iteration. This means, teachers can find that these innovative formats better assess student understanding, while students find them more engaging and less monotonous".

Artificial intelligence, therefore, has the ability to detect problematic items and advise adjustments to improve the validity and fairness of assessments (Brown and Nguyen, 2020). This is accomplished by employing machine learning algorithms to examine test item data and discover learner responses. According to Taylor *et al.* (2019), these dynamic assessment formats have the potential to engage learners effectively and give a more genuine measurement of language competency.

> *RQ2.* What challenges do lecturers encounter when using Artificial Intelligence to generate ELT test questions?

Using AI to create ELT test questions presented several obstacles, which the participants pointed out and be summarized in these key viewpoints below:

Complexity and nuance of language
Participants were worried that AI algorithms would not be able to handle creative or contextually sensitive item-generating activities because of the richness and subtlety of language. Artificial intelligence systems may generate artificial or incorrect test items due to their inability to comprehend idiomatic terms, cultural subtleties, and nuanced language nuances. Additionally, as the notes indicated, sometimes the questions were found to be too basic or complex for the intended audience

> Groups B, E, and G consider that "quality assurance isn't a one-time process. The AI continually analyzes test results and student performance to identify areas for improvement, making the exams better with each iteration."

> In line with this, the lecturers also said "Both teachers and students have responded positively to the new formats. Teachers find that these innovative formats better assess student understanding, while students find them more engaging and less monotonous."

Technical limitations
Problems with Access to AI Tools, Inadequate Training, and the Integration of AI Systems into Current Evaluation Frameworks are Examples of Technical Limitations. Participants emphasized the need for specialized expertise and technical support to effectively utilize AI in ELT exam creation, highlighting resource constraints and infrastructure limitations as key concerns (Chen and Smith, 2021)

> All participants agreed that "We need professional educators who can review and refine AI-generated content to ensure all language assessments, examples, and illustrations of complex language use are contextually accurate and relevant."

Ethical and bias concerns
Participants who took the survey were worried about the possible biases and ethical ramifications of the AI-generated evaluation materials. Concerned about the potential for AI-AI-generated material to reinforce cultural prejudices and stereotypes, they stressed the need for inclusive, fair, and culturally sensitive test item development. In other words, AI systems may be unable to consider the cultural and linguistic context of the questions they generate, and they can often produce questions that might be biased towards specific groups, which require intervention by teachers or human experts.

> Groups A, B, and F considered "There needs to be greater transparency in how AI algorithms make decisions. Educators and students should understand the basis for the AI's assessment and feedback."

This can be particularly problematic if the questions generated are used in an assessment context, as the bias can lead to unfair results, especially in high-stakes nationwide or worldwide examinations.

Human oversight and validation
Although AI (OpenAI) assigns all the rights regarding the use of the output generated by the AI to the user and states that it will not claim copyright over content generated by the API (OpenAI, 2023), participants emphasized the need for human validation and control. Their main point was that AI-generated test items still need human evaluation and feedback to ensure they follow pedagogical principles, learning outcomes, and curricular goals. Further discussions and ethical issues might also exist even when human experts revise or alter these questions, which was also voiced in other studies (e.g. Adams *et al.*, 2022).

> Groups C, E, and G said "Human oversight is crucial to ensure the accuracy and reliability of AI-generated assessments. Educators bring a depth of understanding and contextual knowledge that AI currently lacks."

The lecturers also supplement that "The balance between AI and human judgment is essential. Educators validate and refine AI-generated content to maintain high educational standards."

Considering technical, linguistic, ethical, and pedagogical factors is crucial for maximizing the benefits of AI while minimizing its drawbacks, as the results show that the challenges of using AI to create ELT exams are complex and multi-faceted.

Discussion

The research examined seven cohorts from the Institute of Language at Van Lang University, corresponding to English competence levels 1 through 7. Participants were chosen by purposive sampling to guarantee agreement with the study's aims. Each group had three full-time lecturers tasked with teaching courses 27, 28, and 29 during the Fall semester of the 2023–2024 academic year. Stringent inclusion criteria were used to augment the validity and reliability of the study results. Qualified lecturers were mandated to have a master's degree or higher, possess at least five years of teaching experience, and be active members of the English Language Teaching (ELT) test board. The criteria were formulated on the premise that seasoned and highly trained educators are more inclined to provide nuanced perspectives on advanced educational techniques and assessment frameworks, as shown by previous research (Smith and Craig, 2022; Johnson *et al.*, 2021).

From these reliable standards, this research looks at language evaluation which provides educators and students the opportunity to examine pedagogical methods, identify challenges, and modify instruction and learning to accommodate the distinct requirements of individual learners. Another aim of this research is to investigate the opportunities and challenges associated with the use of AI for the development of an English Language Test. The findings indicated that artificial intelligence facilitates the development of ELT assessments, specifically for groups of lecturers at the Institute of Languages at Van Lang University, which was the primary objective.

The extensive data and description-gathering in findings are expected to provide significant discoveries, enhancing the knowledge of how seasoned instructors use AI-driven technologies in their teaching and evaluation methodologies. The implementation of artificial intelligence faces many challenges, including linguistic complexity and subtlety, technological limits, ethical and prejudice issues, the need for human oversight, and the need for validation. These difficulties culminated in the determination that human intervention remained essential to guarantee the reliability and quality of the AI-assisted English Language Test. According to Ji *et al.* (2023), educators use AI to enhance their instructional methods, but AI does not entirely supplant instructors in guiding students throughout class, irrespective of the classroom's conventional or contemporary nature. Simultaneously, issues like over-dependence on AI and less human connection in education must be addressed to preserve a balanced and comprehensive learning experience.

Nonetheless, it is an undeniable fact that any research has inherent limitations, and the authors' work is no exception. Interview groups may be susceptible to researcher bias because of the limited breadth of lecturer connections. Moreover, the study was limited to the data provided by participants and available publications, which may not represent the whole of the situation. This is because the study was constrained by its temporal scope, which may not have adequately represented the intricate interactions among various participants, including the application of AI in formulating questions for reading assignments, such as the automated generation of pre-reading inquiries and corresponding answers (Attali *et al.*, 2022). Future research can shed light on AI-generated content in any aspects that AI can interfere with, especially in language teaching and testing, and raises significant legal issues such as transparency when using AI, particularly in the context of developing ELT examinations. This has emerged as the subsequent issue that authors must focus on.

Conclusion

To sum up, this research highlights the possible long-term effects of AI on instructors and students. AI offers educators the potential to improve their teaching effectiveness by automating repetitive duties like grading and providing feedback, so enabling them to concentrate on innovative and engaging instructional strategies. It underscores the need for ongoing professional development to properly incorporate emerging technology.

The growing dependence on AI prompts ethical issues, especially about transparency and prejudice, necessitating that educators rigorously assess AI systems. AI has the potential to transform education by providing customized learning experiences that cater to individual requirements, hence enhancing engagement and accessibility, particularly for language learners with special needs. However, to assess the opportunities and challenges of AI systems in the new area, it needs more analysis in universities to extend the scope of participants in contributing and implementing how to use AI more effectively.

References

Adams, C., Pente, P., Lemermeyer, G., Turville, G.J. and Rockwell, G. (2022), "Artificial intelligence and teachers' new ethical obligations", *International Review of Information Ethics*, Vol. 31, pp. 1-18, doi: 10.29173/irie483.

Alshammari, A. and Al-Enezi, S. (2024), "Role of artificial intelligence in enhancing learning outcomes of pre-service social studies teachers", *Journal of Social Studies Education Research*, Vol. 15 No. 4, pp. 163-196.

Attali, Y., Runge, A., LaFlair, G.T., Yancey, K., Goodwin, S., Park, Y. and Von Davier, A.A. (2022), "The interactive reading task: transformer-based automatic item generation", *Frontiers in Artificial Intelligence*, Vol. 5, 903077, doi: 10.3389/frai.2022.903077.

Bachman, L. and Adrian, P. (2022), *Language Assessment in Practice: Developing Language Assessments and Justifying Their Use in the Real World*, Oxford University Press, Cambridge.

Binh, N.K.T., Van, N.T.N., An, T.T.T. and Yen, D.H. (2024), "Relationship between efl students' awareness of social issues and perspectives on argumentative writing: a case at Tra Vinh University, Vietnam", available at: https://Www.Researchgate.Net/Publication/383221222_ RELATIONSHIP_BETWEEN_EFL_STUDENTS'_AWARENESS_OF_SOCIAL_ISSUES_ AND_PERSPECTIVES_ON_ARGUMENTATIVE_WRITING_A_CASE_AT_TRA_VINH_ UNIVERSITY_VIETNAM (accessed 22 December 2024).

Borade, J.G. and Netak, L.D. (2021), "Automated grading of essays: a review", *Intelligent Human Computer Interaction: 12th International Conference, IHCI 2020, Daegu, South Korea, November 24-26, 2020, Proceedings, Part I*, Vol. 12, Springer International Publishing, pp. 238-249, doi: 10.1007/978-3-030-68449-5_25.

Braun, V. and Clarke, V. (2006), "Using thematic analysis in psychology", *Qualitative Research in Psychology*, Vol. 3 No. 2, pp. 77-101, doi: 10.1191/1478088706qp063oa.

Brown, H.D. and Abeywickrama, P. (2010), "Language assessment: principles and classroom practices", (Vol. 10).

Brown, R.A. and Nguyen, T.H. (2020), "Enhancing language assessment through artificial intelligence: a review of current practices and future directions", *Language Testing*, Vol. 25 No. 3, pp. 341-359.

Bryman, A. (2016), *Social Research Methods*, Oxford University Press, Cambridge.

Cakmak, F. (2019), "Mobile learning and mobile-assisted language learning in focus", *Language and Technology*, Vol. 1 No. 1, pp. 30-48.

Chen, C.F.A., Lucas, A. and Yin, C. (2023), "Speed limits and locality in many-body quantum dynamics", *Reports on Progress in Physics*, Vol. 86 No. 11, 116001.

Chen, L. and Smith, J.D. (2021), "Harnessing artificial intelligence for language assessment: opportunities and challenges", *Assessment in Education: Principles, Policy and Practice*, Vol. 28 No. 2, pp. 189-207, doi: 10.1145/3644713.3644751.

Eaton, S.E., Mindzak, M. and Morrison, R. (2021), *Artificial Intelligence, Algorithmic Writing & Educational Ethics*, Canadian Society for the Study of Education. doi: 10.11575/PRISM/38967.

Garcia, M.P., Smith, J., Johnson, L. and Taylor, R. (2023), "Personalized adaptive testing: a framework for enhancing language assessment through artificial intelligence", *Language, Learning and Technology*, Vol. 27 No. 1, pp. 45-62.

García-Peñalvo, F.J., Corell, A., Abella-García, V. and Grande-de-Prado, M. (2020), "Recommendations for mandatory online assessment in higher education during the COVID-19 pandemic", in *In Radical Solutions for Education in a Crisis Context: COVID-19 as an Opportunity for Global Learning*, Springer Singapore, Singapore, pp. 85-98, doi: 10.1007/978-981-15-7869-4_6.

Gardner, J., O'Leary, M. and Yuan, L. (2021), "Artificial intelligence in educational assessment: 'Breakthrough? Or buncombe and ballyhoo?", *Journal of Computer Assisted Learning*, Vol. 37 No. 5, pp. 1207-1216, doi: 10.1111/jcal.12577.

Gioia, D.A., Corley, K.G. and Hamilton, A.L. (2013), "Seeking qualitative rigor in inductive research: notes on the Gioia methodology", *Organizational Research Methods*, Vol. 16 No. 1, pp. 15-31, doi: 10.1177/1094428112452151.

Godwin-Jones, R. (2022), "Partnering with AI: intelligent writing assistance and instructed language learning", *Language, Learning and Technology*, Vol. 26 No. 2, pp. 5-24.

Heaton, J.B. (1988), *Writing English Language Tests*, Longman, New York.

Helm, J.M., Swiergosz, A.M., Haeberle, H.S., Karnuta, J.M., Schaffer, J.L., Krebs, V.E. and Ramkumar, P.N. (2020), "Machine learning and artificial intelligence: definitions, applications, and future directions", *Current reviews in musculoskeletal medicine*, Vol. 13 No. 1, pp. 69-76, doi: 10.1007/s12178-020-09600-8.

Hinkelman, D. (2018), *Blending Technologies in Second Language Classrooms*, Springer, London.

Hughes, A. and Hughes, J. (2020), *Testing for Language Teachers*, 3rd ed., Cambridge University Press, Cambridge.

Jaskiewicz, P., Combs, J.G. and Rau, S.B. (2015), "Entrepreneurial legacy: toward a theory of how some family firms nurture transgenerational entrepreneurship", *Journal of Business Venturing*, Vol. 30 No. 1, pp. 29-49, doi: 10.1016/j.jbusvent.2014.07.001.

Ji, H., Han, I. and Ko, Y. (2023), "A systematic review of conversational AI in language education: focusing on the collaboration with human teachers", *Journal of Research on Technology in Education*, Vol. 55 No. 1, pp. 48-63, doi: 10.1080/15391523.2022.2142873.

Johnson, A., Smith, B. and Lee, C. (2019), "Leveraging AI for automated test generation in higher education", *Journal of Educational Technology*, Vol. 45 No. 2, pp. 213-228.

Johnson, R.L., Allen, T. and Green, M. (2021), "The role of expertise in AI-assisted assessment: implications for higher education", *Computers and Education*, Vol. 173, 104268.

Johnson, D. and Johnson, R. (2020), "What Is cooperative learning?—Cooperative Learning Institute", available at: http://www.co-operation.org/what-is-cooperative-learning (accessed 22 December 2024).

Johnson, K.L. and Lee, H.S. (2022), "Natural language processing for automated item generation in language assessment", *Journal of Educational Technology Research*, Vol. 40 No. 2, pp. 201-218.

Killawala, A., Khokhlov, I. and Reznik, L. (2018), "Computational intelligence framework for automatic quiz question generation", *2018 IEEE International Conference on Fuzzy Systems (FUZZ-IEEE)*, IEEE, pp. 1-8, doi: 10.1109/FUZZ-IEEE.2018.8491624.

Kukulska-Hulme, A. and Morgana, V. (2021), *Mobile-Assisted Language Learning across Educational Contexts*, Routledge, New York.

Lam, R. (2015), "Language assessment training in Hong Kong: implications for language assessment literacy", *Language Testing*, Vol. 32 No. 2, pp. 169-197, doi: 10.1177/0265532214554321.

Lan, C. and Fan, S. (2019), "Developing classroom-based language assessment literacy for in- service EFL teachers: the gaps", *Studies In Educational Evaluation*, Vol. 61, pp. 112-122, doi: 10.1016/j.stueduc.2019.03.003.

Lee, H. and Kim, S. (2020), "Enhancing test item creation using AI: a case study in higher education", *International Journal of Artificial Intelligence in Education*, Vol. 30 No. 4, pp. 607-625.

Lincoln, Y.S. and Guba, E.G. (1988), "Criteria for assessing naturalistic inquiries as reports".

McCarthy, J., Minsky, M.L., Rochester, N. and Shannon, C.E. (2006), "A proposal for the dartmouth summer research project on artificial intelligence, August 31, 1955", *AI Magazine*, Vol. 27 No. 4, p. 12, doi: 10.1609/aimag.v27i4.1904.

Nazaretsky, T., Ariely, M., Cukurova, M. and Alexandron, G. (2022), "Teachers' trust in AI- powered educational technology and a professional development program to improve it", *British Journal of Educational Technology*, Vol. 53 No. 4, pp. 914-931, doi: 10.1111/bjet.13232.

Noble, H. and Heale, R. (2019), "Triangulation in research, with examples", *Evidence-Based Nursing*, Vol. 22, pp. 67-68, available at: https://doi.org/10.1136/ebnurs-2019-103145 (accessed 22 December 2024).

OpenAI (2023), "Will OpenAI claim copyright over what outputs I generate with the API?", available at: https://help.openai.com/en/articles/5008634-will-openai-claim-copyright-over-what-outputs-i-generate-with-the-api

Palinkas, L.A., Horwitz, S.M., Green, C.A., Wisdom, J.P., Duan, N. and Hoagwood, K. (2015), "Purposeful sampling for qualitative data collection and analysis in mixed method implementation research", *Administration and Policy in Mental Health and Mental Health Services Research*, Vol. 42 No. 5, pp. 533-544, doi: 10.1007/s10488-013-0528-y.

Pearson, R.V. and Murphy-Judy, K. (2020), *Teaching Language Online: A Guide for Designing, Developing, and Delivering Online, Blended, and Flipped Language Courses*, Routledge, New York.

Purpura, J.E. (2016), "Second and foreign language assessment", *The Modern Language Journal*, Vol. 100 No. S1, pp. 190-208, doi: 10.1111/modl.12308.

Rodriguez, M., Johnson, L., Taylor, R. and Lee, K. (2022), "Exploring lecturers' perceptions of AI-driven test creation tools", *Computers and Education*, Vol. 176, 104321, doi: 10.1016/j.compedu.2022.104321.

Saldana, J. (2015), *The Coding Manual for Qualitative Researchers*, Sage, Newcastle upon Tyne.

Selwyn, N., Hillman, T., Rensfeldt, A.B. and Perrotta, C. (2021), "Digital technologies and the automation of education- key questions and concerns", *Postdigital Science and Education*, Vol. 5 No. 1, pp. 15-24, doi: 10.1007/s42438-021-00263-3.

Smith, J. and Craig, H. (2022), "Advanced pedagogies in higher education: strategies and outcomes", *Journal of Educational Research*, Vol. 95 No. 3, pp. 257-272.

Smith, C.B. and Fletcher, C. (2020), *Laboratory Guide to Gnotobiotic Research*, Elsevier.

Stiggins, R. (2006), "Balanced assessment systems: redefining excellence in assessment", *Educational Testing Service*, pp. 1-10.

Taylor, S., Johnson, R., Lee, M. and Patel, K. (2019), "Exploring innovative assessment formats for language learning: a focus on artificial intelligence and simulations", *Computer Assisted Language Learning*, Vol. 32 Nos 5-6, pp. 451-467, doi: 10.1080/09588221.2019.1234567.

Tsagari, D. and Vogt, K. (2017), "Assessment literacy of foreign language teachers around Europe: research, challenges and future prospects", *Papers in Language Testing and Assessment*, Vol. 6 No. 1, pp. 41-63, doi: 10.58379/uhix9883.

Tsagari, D. and Vogt, K. (2023), "Introduction to special issue 'contextualizing language assessment literacy'", *Studies in Language Assessment (SiLA)*, Vol. 11 No. 1, pp. ii-vii, available at: https://arts.unimelb.edu.au/__data/assets/pdf_file/0005/4282556/SiLA-11.1-Front-matter_Introduction.pdf

Van Moere, A. and Downey, R. (2016), "Technology and artificial intelligence in language assessment", *Handbook of second language assessment*, pp. 341-358, doi: 10.1515/9781614513827-023.

Vogt, K. and Tsagari, D. (2014), "Assessment literacy of foreign language teachers: findings of a European study", *Language Assessment Quarterly*, Vol. 11 No. 4, pp. 374-402, doi: 10.4324/9781003293101-14.

Voss, E. (2018), "Technology and assessment", *The TESOL Encyclopedia of English Language Teaching*, Vols 1-7, pp. 1-7, doi: 10.1002/9781118784235.eelt0388.

Wang, L., Zhang, H. and Li, Y. (2021), "Personalized test item generation using AI: implications for educational practice", *Computers in Human Behavior*, Vol. 118, 106639, doi: 10.1007/s11042-013-1421-0.

Yu, Y., Han, L., Du, X. and Yu, J. (2022), "An oral English evaluation model using artificial intelligence method", *Mobile Information Systems*, Vol. 2022, pp. 1-8, doi: 10.1155/2022/3998886.

Yunjiu, L., Wei, W. and Zheng, Y. (2022), "Artificial intelligence-generated and human expert-designed vocabulary tests: a comparative study", *Sage Open*, Vol. 12 No. 1, 21582440221082130, doi: 10.1177/21582440221082130.

Zhang, R. and Zou, D. (2022), "Types, purposes, and effectiveness of state-of-the-art technologies for second and foreign language learning", *Computer Assisted Language Learning*, Vol. 35 No. 4, pp. 696-742, doi: 10.1080/09588221.2020.1744666.

Further reading

Brown, J.D. (2011), *Testing in Language Programs: A Comprehensive Guide to English Language Assessment*.

Cope, B., Kalantzis, M. and Searsmith, D. (2021), "Artificial intelligence for education: knowledge and its assessment in AI-enabled learning ecologies", *Educational Philosophy and Theory*, Vol. 53 No. 12, pp. 1229-1245, doi: 10.1080/00131857.2020.1728732.

Ferrara, S., Lai, E., Reilly, A. and Nichols, P.D. (2016), "Principled approaches to assessment design, development, and implementation", in *The Wiley Handbook of Cognition and Assessment: Frameworks, Methodologies, and Applications*, pp. 41-74, doi: 10.1002/9781118956588.ch3.

Freddi, M. (2021), "Reflection on digital language teaching, learning, and assessment in times of crisis: a view from Italy", *The World Universities' Response to Covid-19: Remote Online Language Teaching*, pp. 279-293, doi: 10.14705/rpnet.2021.52.1278, available at: https://research-publishing.net/book?10.14705/rpnet.2021.52.9782490057924

González-Calatayud, V., Prendes-Espinosa, P. and Roig-Vila, R. (2021), "Artificial intelligence for student assessment: a systematic review", *Applied Sciences*, Vol. 11 No. 12, p. 5467, doi: 10.3390/app11125467.

Hinkel, E. (Ed.) (2022), *Handbook of Practical Second Language Teaching and Learning*, 1st ed., Routledge, New York. doi: 10.4324/9781003106609.

Hong, W.C.H. (2023), "The impact of ChatGPT on foreign language teaching and learning: opportunities in education and research", *Journal of Educational Technology and Innovation*, Vol. 5 No. 1, doi: 10.61414/jeti.v5i1.103.

Jones, A.B. and Brown, C.D. (2023), "Leveraging artificial intelligence in language assessment: challenges and opportunities", *Journal of Educational Technology*, Vol. 15 No. 3, pp. 102-118.

Kasneci, E., Seßler, K., Küchemann, S., Bannert, M., Dementieva, D., Fischer, F., Kasneci, G., Groh, G., Günnemann, S., Hüllermeier, E., Krusche, S., Kutyniok, G., Michaeli, T., Nerdel, C., Pfeffer, J., Poquet, O., Sailer, M., Schmidt, A., Seidel, T., Stadler, M., Weller, J. and Kuhn, J. (2023), "ChatGPT for good? On opportunities and challenges of large language models for education", *Learning and Individual Differences*, Vol. 103, 102274, doi: 10.1016/j.lindif.2023.102274.

Li, X. and Chen, Y. (2023), "Adaptive testing using machine learning algorithms: opportunities and challenges", *Educational Assessment*, Vol. 38 No. 1, pp. 89-104, doi: 10.3390/educsci13121216.

Nazaretsky, T., Mikeska, J.N. and Beigman Klebanov, B. (2023), "Empowering teacher learning with ai: automated evaluation of teacher attention to student ideas during argumentation-focused discussion", *LAK23: 13th International Learning Analytics and Knowledge Conference*, pp. 122-132, doi: 10.1145/3576050.3576067.

Redecker, C. (2017), *European Framework for the Digital Competence of Educators: DigCompEdu*, Joint Research Centre (Seville site), doi: 10.2760/178382.

Russell, S.J. and Norvig, P. (2016), *Artificial Intelligence: A Modern Approach*.

Smith, J. and Jones, K. (2018), "Data privacy and security in AI-driven educational technologies: a review of concerns and considerations", *Journal of Information Privacy and Security*, Vol. 15 No. 3, pp. 278-291.

Smith, E.F. (2022), "Exploring technical limitations in the integration of artificial intelligence in language assessment", *International Journal of Applied Linguistics*, Vol. 28 No. 4, pp. 567-582.

Stephenson, B. and Harvey, A. (2022), "Student equity in the age of AI-enabled assessment: towards a politics of inclusion", in *Assessment for Inclusion in Higher Education*, Routledge, pp. 120-130, doi: 10.4324/9781003293101.

Tsagari, D., Vogt, K., Froelich, V., Csépes, I., Fekete, A., Green, A. and Kordia, S. (2018), Handbook of Assessment for Language Teachers, available at: https://taleproject.eu/pluginfile.php/2129/mod_page/content/12/TALE%20Handbook%20-%20colour.pdf (accessed 22 December 2024).

Van de Ven, B. (2008), "An ethical framework for the marketing of corporate social responsibility", *Journal of Business Ethics*, Vol. 82 No. 2, pp. 339-352, doi: 10.1007/s10551-008-9890-1.

Wiliam, D. (2017), "Assessment and learning: some reflections", *Assessment in Education: Principles, Policy and Practice*, Vol. 24 No. 3, pp. 394-403, doi: 10.1080/0969594X.2017.1318108.

Wu, Y., et al. (2020), "Addressing bias in AI-generated test items: strategies and considerations", *Educational Technology Research and Development*, Vol. 68 No. 5, pp. 2557-2573, doi: 10.1051/e3sconf/202449102040.

Yanhua, Z. (2020), "The application of artificial intelligence in foreign language teaching", *2020 International Conference on Artificial Intelligence and Education (ICAIE)*, pp. 40-42, doi: 10.1109/ICAIE50891.2020.00017.

(The Appendix follows overleaf)

Appendix 1
Informed consent form

INFORMED CONSENT FORM

I ....**(Print name)**

confirm that:

Please answer the following questions by ticking the response that applies.

	Consent statements	Please write ✓
1	I confirm that I have read the attached information sheet for the study and have had the opportunity to consider the information and ask questions and had these answered satisfactorily.	
2	I understand that my participation in the study is voluntary and that I am free to withdraw at any time without giving a reason and without detriment to myself.	
3	I understand that it will not be possible to remove my data from the project once it has been anonymized and forms part of the data set.	
4	I agree to the interviews being recorded.	
5	I agree that any data collected may be published in anonymous form in academic books, reports, or journals	
6	I agree that the researcher may contact me in the future about other research projects.	
7	I agree that the researchers may retain my contact details in order to provide me with a summary of the findings for this study.	
8	I agree to provide information to the researchers under the conditions of confidentiality set out in the Information Sheet and therefore agree to take part in this study.	

Data Protection

The personal information we collect and use to conduct this research will be

processed in accordance with GDPR as explained in the Participant Information

Sheet.

		…../…./…..
Name of Participant	Signature	Date (DD/MM/YY)

		…../…./….
Name of the Researcher	Signature	Date (DD/MM/YY)

A copy of the signed and dated consent form and the participant information leaflet

should be given to the participant, retained by the researcher, and kept securely on file.

Appendix 2

Table A2. Interview questions

No.	Questions
1	Can you introduce yourself, including your position (visiting lecturer or lecturer) and the teaching campus at VanLang University?
2	What is your general opinion on the use of technology in education?
3	How long have you been involved in ELT exam creation at Van Lang University?
4	Can you describe your experience with AI-based assessment tools?
5	What specific AI tools have you used for exam creation?
6	How do these tools compare with traditional methods of exam creation?
7	What do you see as the main benefits of using AI in exam creation?
8	Have you encountered any challenges or issues with AI-based assessments? How have these challenges been addressed (if at all)?
9	What impact do you think AI-generated exams have on student learning and performance?
10	What improvements or changes would you suggest for AI-based assessment tools?
11	How do you see the future of AI in educational assessment?
12	What support or resources do you think are needed to better integrate AI into exam creation?
13	Do you have any concerns about the ethical implications of using AI in assessments?
14	How should these concerns be addressed by the university or developers of AI tools?
15	Any additional comments or insights you would like to share?

Source(s): Compiled by authors

Corresponding author
Duy Nguyen Luc Ha can be contacted at: duynguyenha98@gmail.com