https://doi.org/10.15625/2615-8965/22610108

Pedagogical Challenges of Formative E-Assessment in Vietnam's Tertiary Education

Doan Kim Khoa[1], Bui Do Cong Thanh[2]✉, Le Van Thinh[3], Nguyen Thi Xuan Lan[4], Tran Le Nghi Tran[5]

[1] khoa.dk@ou.edu.vn
Centre for Educational Research and Innovation, Ho Chi Minh City Open University
97 Vo Van Tan, Xuan Hoa Ward, Ho Chi Minh City (Vietnam)
Scientific positions: M.Ed
ORCID: https://orcid.org/0009-0008-5916-7805

[2] thanh.bdc@ou.edu.vn
Faculty of Foreign Languages
Ho Chi Minh City Open University
35-37 Ho Hao Hon, Cau Ong Lanh Ward, Ho Chi Minh City (Vietnam)
Scientific position: M.Ed
ORCID: https://orcid.org/0009-0005-8948-4041

[3] lethinhpy@yahoo.com
Department of Foundation Training,
Banking Academy of Vietnam, Phu Yen Campus, (Vietnam)
441 Nguyen Hue, Tuy Hoa Ward (Vietnam)
Scientific position: PhD
ORCID: https://orcid.org/0000-0001-6856-9883

[4] lan.ntx@ou.edu.vn
Centre for Educational Research and Innovation, Ho Chi Minh City Open University
97 Vo Van Tan, Xuan Hoa Ward, Ho Chi Minh City, (Vietnam)
Scientific position: PhD.
ORCID: https://orcid.org/0009-0005-6579-2226

[5] tran.tln@ou.edu.vn
Centre for Educational Research and Innovation, Ho Chi Minh City Open University
97 Vo Van Tan, Xuan Hoa Ward, Ho Chi Minh City, (Vietnam)
Scientific position: PhD
ORCID: https://orcid.org/0000-0001-6526-1269

✉ Corresponding author

ABSTRACT: *Recently, the adoption of e-learning in tertiary education in Vietnam has been phenomenal, necessitating a critical investigation into e-assessment practices. Across the country, the prevalence and diversity of formative e-assessment (FE) have raised quality concerns among teachers. This study, therefore, identifies the pedagogical challenges and proposes some practical solutions to maintain the quality of FE following the principles of validity, reliability, authenticity, practicality, and washback. Qualitative data were gathered through in-depth interviews with thirteen teachers and open-ended questionnaire responses from 103 lecturers across various institutions. Thematic analysis revealed major challenges including students' lack of academic integrity, over-reliance on AI-enhanced tools, the nature of test task design, types of assignments, teachers' lack of institutional guidance, teachers' delayed feedback, time pressure for students, and varied levels of contributions in groupwork. In response to these pedagogical challenges, lecturers insisted on the need for professional development in FE design, proctoring, grading, and giving feedback. There were also calls for alternatives to conventional FE assignments, relevant and meaningful task design to meet the demands of the real world while informing students' proper attitudes towards AI use.*

KEYWORDS: **e-assessment, e-learning, formative e-assessment (FE), FE design, tertiary education, assessment principles.**

➔ Received 25/11/2025 ➔ Revised manuscript received 4/3/2026 ➔ Published 12/5/2026.

1. Introduction

The post-COVID-19 period marked an acceleration of online learning worldwide, particularly e-learning. In 2021, the global e-learning market was valued at US$214.26 billion, with an expected annual growth of 20.5% until 2030 (Polaris Market Research Analysis, 2022). In line with the rapid development of e-learning, technologies for testing and assessment have also seen significant growth. A variety of ICT tools and pedagogical strategies have been developed to boost the effectiveness of e-learning and e-assessment. However, recent studies have identified various challenges regarding e-assessment (Wahas & Syed, 2024), highlighting the need for further research to better understand and address these issues.

2. Literature Review

2.1. E-learning and E-assessment

The term "e-learning", a shortened form of "electronic learning", was first used in 1998. It often refers to a technology-assisted learning system known as web-based learning or Internet-based learning (Aulakh *et al.,* 2023). The precise definitions of e-learning over the past two decades might vary depending on different

academic views; however, it is generally a form of distance education for remote learners. In reality, e-learning is often used interchangeably with online learning, which is associated with a synchronous or asynchronous online learning environment through the use of the Internet regardless of students' physical location (Singh & Thurman, 2019).

In e-learning environments, e-assessment is one of the key components that contribute to its effectiveness. E-assessment is also known as online, digital, computer-based or web-based assessment (Orsi & Juliano, 2021). From a pedagogical perspective, it is a continuous process used "to measure, monitor, and improve learning, degree of achievements, outcomes, and decide to what extent objectives are accomplished" (Bin Mubayrik, 2020, p. 2). Similar to conventional assessments, e-assessment serves both formative and summative functions. More specifically, summative assessment is often a final graded achievement at the end of a unit or course, while formative assessment is conducted multiple times during the process and targets students' learning improvement based on meaningful feedback (Yambi, 2018).

2.2. Formative E-assessment and Its Usefulness

Compared with face-to-face assessment, e-assessment often involves asynchronous interactivity among online participants (both teachers and learners). E-assessment activities, when based on the formative and continuous perspective, enable students to take an active role in their learning (Romeu Fontanillas *et al*., 2016). One common type of formative e-assessment (FE) is an online multiple-choice test, which was found to improve knowledge gain, exam performance, confidence, learner satisfaction, metacognition, and self-regulatory behaviours (Say *et al*., 2022). Other FE tasks, such as e-portfolios and Moodle quizzes, can be both product- and process-oriented and employ adaptive techniques using artificial intelligence (AI)-supported feedback to reduce teachers' workload (Gün-tosik *et al*., 2023). E-assessment can be both consistent and reliable while allowing lecturers to modify an activity based on the learners' needs (Blanco & Ginovart, 2012).

According to a systematic review by Lei and Lei (2026), an increasing number of studies, particularly in Asian contexts (e.g., China, Iran, Indonesia) indicate formative assessment and teacher assessment literacy have become educational policy priorities. In the context of Vietnam, some studies reveal that students have generally positive attitudes towards FE despite some obstacles (Trinh & Trinh, 2023), and that FE has significant impacts on writing achievement (Nhu & Tin, 2019). Therefore, it is recommended that teachers should employ FE to enhance English language teaching and testing.

2.3. Advantages and Disadvantages of Formative E-assessment in Tertiary Education

Compared to the traditional pen and paper method, FE has some advantages, such as better interaction and greater personalization, more authentic and engaging tasks (Orsi & Juliano, 2021). Improved behavioural outcomes and evaluation of the students were observed in various studies conducted in different undergraduate courses (Orsi & Juliano, 2021; Say *et al*., 2022; Blanco & Ginovart, 2012). However, it was also found that different FE designs can influence students' level of engagement, attitudes, and academic performance (Chen & Chen, 2023) due to the complexity of tasks. For example, Prendes-Espinosa *et al*. (2022) pointed out that both teachers and students struggle with the time involved in FE development. Other disadvantages included the amount of effort required for teacher feedback, grading moderation and alignment with peer or self-evaluation.

In addition, the rapid development of AI used in education has a significant impact on formative e-assessment, bringing both benefits and constraints. On the one hand, AI can assist educators in enhancing assessment efficiency, and developing feedback mechanisms. Several studies conducted on automated grading or automated essay scoring systems in undergraduate courses show an agreement between computer and human raters, which could reduce cost and time (Zawacki-Richter *et al.*, 2019). As reported by Chen *et al*. (2020), different AI applications

can support instructors by providing "built functionalities" to evaluate performance, grade, and provide feedback to students (p. 75273). On the other hand, some ethical concerns regarding educational integrity have emerged whereas educators seem to have insufficient preparation for tackling such issues. For instance, students can use a chatbot application to generate high-quality written assignments like essays, which could cause inequities in assessment (Cotton *et al*., 2024). In one study by Ma *et al*. (2025), most Chinese in-service teachers, although acknowledging the importance of teaching AI ethics, lacked expertise in this area because it was not adequately addressed in their prior education or professional development activities. Meanwhile, there is an alarmingly increasing number of higher education students using AI tools to cheat in exams, and the cheating behaviours have constantly changed (Nartgün & Kennedy, 2024). Therefore, it is essential to better equip educators with FE strategies.

2.4. Key Principles of Formative E-assessment in Tertiary Education

There have been various studies examining the key principles in implementing formative e-assessment in higher education contexts (De Villiers *et al*., 2016; Padayachee *et al*., 2018; Jamalludin & Romli, 2023). These projects generally covered student experience and technological matters. However, while these studies helped further understanding into the accuracy of e-assessment tools, they often overlooked the critical pedagogical aspect of FE. The current study, therefore, adopts the five key principles of assessment proposed by Brown and Abeywickrama (2019) to consider FE consistently in terms of practicality, validity, reliability, authenticity, and washback. Practicality is related to logistical and administrative matters (i.e., time and scoring). Validity focuses on what tests are intended to measure while reliability centres on the consistency of test results across different times and students. Authenticity refers to how tests reflect real-life situations. Lastly, washback is about the test effect on learning and teaching processes.

In a recent systematic review, Pordanjani and Salehi (2025) noted the growing prominence of e-assessment and a range of challenges that affect its reliability, equity, and pedagogical value, including technical issues, academic integrity, data privacy, and marking difficulties. The articles published between 2000 and 2024, however, focused mostly on e-assessment in general (Pordanjani & Salehi, 2025), with formative e-assessment being largely overlooked and under-represented in the literature. Similarly, in the context of Vietnam's higher education, formative e-assessment, as an emerging area of interest, is yet to be regulated by the policies of the Vietnam Ministry of Education and Training. Circular No 28/2023/TT-BGDĐT, issued by the Ministry of Education and Training (2023), required final examinations, which are categorized as one type of summative assessment, to be conducted offline, leaving formative e-assessment in limbo as there is no clear regulation or policy for it. Lecturers and universities, therefore, have the flexibility of choices for what to do and how to do it, which can be both empowering and problematic (Dave, 2025). Considering the important role of formative e-assessment in informing teachers' teaching and students' learning processes (Pacheler *et al*., 2010), it is crucial to further understand formative assessment and provide research evidence for future policy development regarding FE in the context of Vietnamese higher education. The current study, therefore, aimed to answer the following questions:

- What pedagogical challenges have teachers encountered in conducting formative e-assessments in Vietnamese universities?
- What actions have been taken by these teachers to address the challenges identified?

3. Methodology

3.1. Research Design

This study deployed a qualitative explanatory research design to gain an in-depth and thorough understanding (Creswell & Creswell, 2018) of pedagogical challenges that lecturers faced when conducting FE. In addition, pedagogical solutions to the existing challenges were proposed to ensure the validity, reliability, authenticity, practicality,

and washback of the FE implemented in the e-learning settings. Semi-structured interviews, compared to a pre-set guide of structured interviews, can make a fuller advantage of the knowledge that can emerge through the dialogues between the interviewer and interviewee (Brinkmann, 2018). The research is primarily based on the qualitative data collected from the individual semi-structured interviews. For data triangulation, the responses from an open-ended question of a prior online survey were also used as an overall pre-interview input for the researchers before starting the interviews. These responses, as stated by Fraser (2024), are appropriate for describing participants' viewpoints to explore the common norms.

3.2. Participants

The purposive sampling method was adopted, calling for the active and voluntary participation of 13 lecturers (eight females, five males) among 103 lecturers who responded to the survey conducted prior to the interview. Purposive sampling helps the researchers to apply "intentionality in whom they recruit to participate" in the qualitative interviews (Corwin & Harper, 2024). It ensured that the participants came from different faculties (i.e., English Linguistics, Business Administration, Business and Law, and Civil Engineering) of higher institutions across Vietnam. The researchers intentionally invited those who were competent and experienced enough to answer questions pertaining to the subject matter (Mulisa, 2022), since they had experience in teaching and testing online. The interviewees remained anonymous with their names encoded from A to M. These lecturers had teaching experience for bachelor's degree online or e-learning programs for at least five years. Besides teaching, they also partook in the curriculum design and course development for these online university degree programs.

3.3. Data Collection and Analysis

The data for this study were collected as part of research project with funding from Vietnam's Ministry of Education and Training in two rounds. In the first round, an online survey was sent via email to lecturers who taught online undergraduate programs in universities across the country. A total of 103 teachers responded to the survey, but three teachers did not teach online. Therefore, the data collected from 100 teachers provided insights into their beliefs and practices in online formative assessment. Data from the survey helped identify the common challenges of formative e-assessments encountered by teachers across Vietnamese universities. The survey included open-ended questions as it was important to focus on explaining how the participants addressed their main concerns, rather than just exploring them (Christodoulou, 2025). A *small q* approach, as stated by Fraser (2024), was used to code the open-ended data into numerical data for descriptive statistical analysis. Each response was categorized into the common themes regarding the challenges in FE, and a simple numeric analysis was conducted to summarize the information. The survey responses helped inform the questions used in the interviews.

At the end of the survey, interested teachers were invited to attend an online interview to further discuss the challenges encountered and solutions they implemented for formative e-assessment. Among the survey respondents who expressed interest in a follow-up interview, twenty were invited to participate based on the above selection criteria, and thirteen accepted the invitation. In the second round, thirteen in-depth interviews were conducted online via Zoom in the Vietnamese language. Each interview was conducted in 45 minutes, recorded using two devices, transcribed non-verbatim, and stored as soft copies in two password-protected computers before analysis. Overall, the key criteria for selecting the participants included minimum years of relevant teaching experience, different majors, different universities based in different regions of the country. These selection criteria ensure the demographic diversity of the participants in terms of experience, background, and context.

Both the qualitative survey and interview data were coded and analysed based on the five principles regarding the pedagogical aspects of

FE (*validity, reliability, authenticity, practicality, and washback*) (Brown & Abeywickrama, 2019). Thematic analysis was used to identify and interpret patterns or themes after identifying keywords, codes, and themes in the data set (Naeem *et al*., 2023). Two research team members read the transcripts several times to familiarise themselves with the content before coding them independently. The coding process was repeated several times until no new theme was identified. Then, the two coders compared their coding results and discussed the differences before entering the finalised codebook into Nvivo. The final codes were organised into themes, with retrievable coded data for review by other members of the research team. The iterative independent coding and checking process ensures the reliability, transparency and rigor of the analysis process.

To facilitate teachers' understanding of the survey and interview questions, Vietnamese was used for data collection. The research team then translated all the responses into English and had two lecturers in charge of Translation courses to check the translation to maintain faithfulness. The interviewees were also given the opportunity to check the transcripts and their English translation before the analysis started.

3.4. Research Trustworthiness

The trustworthiness of qualitative research involves credibility, transferability, dependability, and confirmability, which are generally comparable to validity and reliability in quantitative research (Denzin, 2014). The qualitative data in the current study were collected from the in-depth interviews, with data triangulation ensured through open-ended questions via a survey. The fact that the survey responses were screened from 103 to 100, along with the selection criteria for the interview participants, indicates the steps to ensure the reliability of the data. This helps enhance the research's credibility. For data analysis, two members of the research team were in charge of reviewing the themes and interpretations of the data to confirm its precision, increasing the dependability of the study. To reduce the risk of bias, the researchers regularly cross-checked and confirmed the themes that emerged from the data with all participants to ensure accuracy. All these measures contributed to the trustworthiness of the research.

3.5. Ethical Considerations

Invitation emails were sent to the target survey respondents and interviewees to seek their voluntary participation in the ten-minute survey or the forty-five-minute interview. Upon their agreement, the consent forms were sent to them as well, confirming that the information they provided was kept entirely confidential and intended for research purposes only. The researchers clearly explained the research objectives, the nature of the data collection, and the use of the collected data. Importantly, anonymity and confidentiality were highly maintained in order to protect the participants' identities, prevent bias, and avoid conflicts of interest. The participants were also informed that they could withdraw from the research at any time and for any reason. The communication during the interviews was open, honest, and transparent. The interviewees could feel free to end the talk at their will without any consequences.

4. Results

The following sections present part of the data collected for the research project to answer the research question on pedagogical challenges and solutions to address them regarding formative e-assessment.

4.1. Survey Open-ended Responses

With regard to FE, the current study identifies a range of pedagogical challenges to the implementation of formative e-assessment as shown in Table 1.

Table 1 summarizes the eleven common obstacles reported by 100 lecturers in various Vietnamese public and private universities. Among these, four most common challenges related to students' conduct, followed by five teacher-related issues, with the two least popular relating to the institute level. They were synthesized and categorized in terms of repeated

Table 1. A summary of the Common Pedagogical Challenges Reported by Survey Respondents (n=100)

No	Percentage	Challenge
1	83%	Students ask others to do assignments for them.
2	71%	It is difficult to detect students using AI.
3	71%	It is difficult to monitor students during assignment completion.
4	69%	Students plagiarize from others.
5	48%	It is time-consuming to grade assignments.
6	37%	It takes much effort to give detailed feedback on individuals.
7	27%	Teachers lack professional development activities regarding e-assessment.
8	21%	Teachers lack e-assessment skills.
9	18%	It is difficult to adjust assessment formats from offline to online.
10	17%	There is a lack of detailed institutional policies or guidelines.
11	11%	There is a lack of specific instructions from the faculty level.

patterns of the relevant themes. The most noticeable concern is about academic integrity, with 83% of the lecturers reporting students engaged in contract cheating, 71% noting a sign of using AI, and 69% mentioning plagiarism. 71% of the lecturers having difficulty detecting academic misconduct or managing academic integrity in online environments. Additionally, time and effort in implementing FE and giving feedback were also quite common among the lecturers (48% and 37%, respectively). Furthermore, there is a need for professional support, as approximately one-fifth of the lecturers indicated the need for professional development in e-assessment training, so that they will not simply replicate conventional test formats but adapt them effectively.

4.2. Interview Responses

4.2.1 Validity

In synchronous FE, validity, which measures what is intended to be measured, is greatly challenged in digital environments by external interference and misaligned task design. For external factors such as technology use, ten out of thirteen interviewed lecturers stated that their students turned to translation tools or online dictionaries to assist them in producing answers, thereby leading to an increase in the level of accuracy beyond their lecturers' expectations. For example, one lecturer reflected:

I tried to have the lockdown browser installed on the web page. Once any of the students switched to another external resource, the exam web page automatically closed down ... However, I accidentally overheard students' conversation during the non-test time about the group chat they created to support one another [to get answers during the assessment]. (Lecturer A)

Students' cheating makes their test results unreliable, but the above reflection also raises concerns regarding how to ensure the validity of the assessment tasks so that lecturers can measure the test constructs they intended to regardless of students' use of technology. For example, for asynchronous e-assessment tasks, all of the interviewees complained about students' use of AI tools and felt doubtful whether they were marking students' own submissions or AI-generated work. If the latter was the case, the assessment became invalid as it failed to evaluate students' skills and knowledge. Even a lecturer with some prior experience in testing and assessment still had difficulty in maintaining FE validity:

... When given assignments, students turn to AI for support, sacrificing their own intellectual investment. I dare not say all of them, but the

majority of them will choose the most convenient way to complete their assignments [which is to use AI tools]. (Lecturer M)

Besides technology use, assessment design also impacted the validity of FE. For example, eight out of thirteen lecturers revealed that the common test task designed to measure critical thinking skills, production, or communication was mainly multiple-choice questions as one of them reported:

Students got high grades for their answers in the weekly assignments that had been designed in the form of multiple choices. However, during the synchronous online conference, they failed to answer my questions. I am sceptical about the selected test task and whether it actually measured what was intended to be measured. (Lecturer E)

Apparently, the weekly assignments using multiple choice questions failed to assess the students' target ability, which hurt the validity of the test. This was confirmed by the later synchronous online conference, which suggested that teachers may also use technology to ensure the test-task validity by choosing the appropriate task design.

4.2.2. Reliability

In FE, reliability refers to consistency, fairness, and dependability of assessment outcomes, all of which are closely linked to teaching and testing methods, assessment design, and human factors (Sarigoz, 2023). Challenges found in the data set are, therefore, categorised into institutional inconsistency, the complexities of group-based assessment, and affective mediators.

In terms of institutional inconsistency, the participants noted the lack of unanimous institutional guidance on the test design. Without standardised protocols, five lecturers shared that they developed varied test formats, and difficulty levels for the same subject. In addition, there is a case in which one lecturer (Lecturer G) confessed that they used the same test for his/her three different classes at different time slots. Consequently, students taking the test first kept their remaining peers informed of the questions that might be reused by their lecturer. This significant spillover effect was observed in synchronous FE, highlighting the significance of ensuring reliability in FE which requires continuous modification based on student feedback and test-item analysis (Blanco & Ginovart, 2012).

With regard to FE via groupwork, especially for asynchronous mode, there were several reliability issues. The findings highlight three specific "reliability gaps": contribution variance, rubric limitations, and academic integrity. Lecturers struggled to ensure that grades reflected individual effort as equal marking for unequal contributions violates the principle of assessment fairness (Laato & Murtonen, 2020). In addition, despite the presence of rubrics, students' interpretations may be varied when it comes to peer assessment. Also, the complexity of the rubrics often failed to match the task requirements, leading to student scepticism regarding the scores allocated.

My students are not eager for groupwork, commenting that the level of peer involvement varies. When the teacher grades group work, despite the availability of rubrics, students still question the reliability of scores allocated for each of them. (Lecturer I)

Another issue that is of great importance related to group work is plagiarism (Prendes-Espinosa *et al.*, 2021). When a particular group member commits plagiarism for his or her assigned part, others are serious about the originality of their work. So how to mark the overall group performance poses a big challenge for the teacher. Two of the lecturers (Lecturers E and L) shared that they had to ask their students to resubmit the paper, adding that most of the students in the group insisted on their ethical commitment to their parts. It may be unfair for those who truly demonstrate their academic ethics while some of their peers did not. This may pose a serious question as to how to ensure reliability in group work for asynchronous e-assessment. Like validity, ensuring reliability in FE is a multifaceted challenge in that it involves the consistency of not only assessment results but also the assessment performances across technological platforms (Mohammed & Mashkhal, 2023).

Referring to affective mediators, in both synchronous and asynchronous online settings, affective factors such as anxiety, inhibition, and motivation together with timing pressure and surroundings (i.e., home distractions, background noise) are very likely to affect performance. In other words, this challenge refers to synchronous pressure and asynchronous procrastination (Padayachee *et al*., 2018). Eight lecturers shared the same story that their students feel truly pressured when dealing with the test during the synchronous online meeting and that some of them did not have enough motivation to keep track of all assignments timely. As a result, they delayed completing the assignments until the deadline loomed. One lecturer reported that *"Most of my students said that they performed more terribly when it comes to synchronous assessment than asynchronous assessment, which inherently allowed them more time to prepare their answers"* (Lecturer A). However, another issue arises when asynchronous assessment tasks were adopted, as recalled by another lecturer:

"Some of the students often completed their assignments in a rush, even when the due date was around the corner. Though they understood that dealing with assignments earlier gave them more time to research and revise their work. Maybe procrastination was the result of low motivation." (Lecturer D)

Obviously, rushed submission may not reflect the true performance that students were capable of, which in turn impacted the reliability of the assessment task. These results revealed that reliability concerns regarding FE were the most common and serious challenges reported by the lecturers.

4.2.3. Authenticity

The findings related to authenticity revealed pedagogical reductionism in skill-based subjects and the physicality gap in experimental learning. For any e-assessment pertaining to experiments or skill practice, the question design was over-reliant on multiple-choice questions, meaning that memorization or rote learning is required (McCallum & Milner, 2020). Similarly, skill practice in foreign language learning is not limited to objective testing design; in fact, subjective testing such as essays, project-based learning, and speaking practices is more sensible. Even if subjective testing (i.e., essays, reports) is adopted and real-life situations are incorporated into the test design to make it look more authentic, students do not really learn much from this kind of assessment, especially when it comes to experiments.

Lecturers, when dealing with e-assessment in experiments, come to terms with the fact that students may reach the expected outcome despite their lack of exposure to real-life experiments. This can be considered a gap in e-learning, especially FE as experiments demand direct exposure. (Lecturer C)

Lecturer C expressed their concern that experiment-related tests are hard to do because their students cannot be exposed to real-life situations, not to mention required equipment or others for performing a task. In other words, the lack of authenticity among the assessment tasks highlighted the risk that students may not be able to perform real-life tasks they will encounter in their professional life and workplace upon graduation.

4.2.4. Washback

In students' terms, for synchronous e-assessment, they have to deal with task completion under time pressure with more focus on speed and test-taking strategies rather than a deep understanding of the subject matter. Meanwhile, for asynchronous mode, the auto-graded multiple-choice task design enables students to prioritize memorisation over communicative skills and other higher-order thinking skills, namely critical thinking, creative thinking, analytical thinking and so forth. The two following excerpts present the lecturers' concerns that the test tasks used had little positive washback for students' professional practice:

It seems that my students successfully earned high scores when dealing with multiple-choice questions whilst they could not write or speak in the academic environment. (Lecturer E)

Most of my students could not draft even a simple blueprint for a particular civil works

project. They selected the correct answers but failed to put theories they had learnt into a masterplan. (Lecturer C)

On the other hand, giving timely feedback presents a challenge as teachers have to deal with a large number of students in addition to time pressure. Students expect to receive feedback for improvement during both synchronous and asynchronous e-assessment. In fact, they obtain feedback, but due to the low quality of feedback-giving, students do not gain much from their teachers' comments.

Being in charge of a large-size class, I could not manage to give feedback to all students. And if yes, I would offer feedback to students who made serious mistakes. Though I know feedback quality is crucial, I could only give feedback at the superficial level. At times, I feel sorry for my students. (Lecturer F)

The above lecturer admitted that they could offer little quality feedback for students' work, resulting in negative washback for students' performance as they may not know what to do and what to avoid in performing similar tasks in the future.

4.2.5. Practicality

In the interviews, all the lecturers mentioned their main challenge regarding time. FE consumes considerable time from designing and adjusting tasks (Lecturers K, G, and D) to grading and giving feedback to individuals or groups (Lecturers F and I). However, the lecturers seemed to accept that there was no perfect assessment task and prioritised practicality more than other principles in their FE practice. For example, Lecturer K believed their current FE practices offered a high level of practicality for not only teachers but also students while understanding that the other principles such as validity and authenticity could be negatively impacted:

For FE, I could not think of any easier ways to design tasks and save grading time. I know objective testing fails to measure communicative skills or critical thinking skills. But at least I find this kind of assessment rather familiar and useful to measure students' knowledge of the subject matter. (Lecturer K)

The lecturer, while understanding the strengths and weaknesses of multiple-choice questions, insisted on the importance of practicality, which drove their decision in keeping them in place.

In summary, the results revealed lecturers' major concerns regarding validity and reliability in terms of academic integrity, affective factors, the students' level of involvement in groupwork, FE design, and needs for professional support. However, many chose to adopt the conventional FE designs like multiple-choice questions thanks to their practicality while understanding that low-quality feedback negatively impacted the authenticity and washback of the assessment tasks.

5. Discussion of Findings

Rather than simply identifying multiple problems in FE, these findings highlight that concerns about reliability and validity stem from deeper systemic issues, including academic integrity, group work assessment, assessment design, and student-related affective factors. This suggests that such challenges are not isolated, but are structurally embedded within current e-assessment practices. For validity, students may ask others, use external resources or turn to AI for help. Such students' use of technology made the assessment invalid and unreliable due to their inflated performance, which aligns with the findings by Cotton *et al*. (2024), reinforcing the argument that technological affordances, while beneficial, simultaneously pose new threats to assessment integrity that institutions have yet to effectively mitigate. Concerning groupwork, lecturers find it extremely hard to grade groupwork fairly as some of the students may be more committed than others. The guarantee among team members regarding plagiarism is in doubt, too. Despite the availability of rubrics, their practical effectiveness appears limited, indicating that assessment tools alone are insufficient without complementary mechanisms to ensure accountability and fairness, a concern similarly raised by Yan *et al*. (2025).

Other challenges related to validity and reliability that were identified in the interview data set refer to the discrepancies among

lecturers in test design and level of difficulty. These inconsistencies underscore a broader lack of standardisation across lecturers, which is consistent with the results by Kearns (2012) on the consequential inconsistency and inaccuracies in grading. Moreover, using the same test for different test times may result in unfairness when students taking the test later gain more academic advantages. This finding suggests the real need for professional support from the institutional level in terms of training and testing policies. Meanwhile, from the perspective of the test-takers, affective factors namely anxiety, inhibition, motivation and personality (i.e., procrastination) also put reliability in jeopardy, confirming the results by Padayachee *et al*. (2018). In this respect, students have to meet tight deadlines, reducing their chances of getting high scores in asynchronous FE due to time pressure and lack of in-depth research and thorough preparation to complete their assignments before the due date. Similarly, their performance in synchronous FE is not as good as that in asynchronous FE on account of real-time academic pressure, nervousness and lack of confidence as confirmed in the findings by Hung *et al*. (2024).

The principle of authenticity is challenged when digital tools fail to replicate real-world contexts or tasks, especially for laboratory experiments and language speaking tasks. This suggests a persistent failure to achieve authenticity in digital assessment, particularly in contexts requiring practical competence and higher-order thinking skills whilst assessment remains dominated by lower-order multiple-choice formats, indicating a clear misalignment with expected learning outcomes. This test task design reflects more than convenience, aligning with the efficiency-driven tendencies noted by Hai (2022) and points to a narrowing of pedagogical priorities. Consequently, such practices are prone to producing negative washback, promoting "teach-to-test" approaches that dent students' motivation and higher-order thinking skills development. This concern echoes Christiana (2025), suggesting a broader and more critical systemic issue in digital assessment design.

Noticeably, the problem with practicality is explicitly shown through massive practical barriers and demands of assessment directly pertaining to untimely and less elaborate feedback. While accepting that there would be no perfect assessment tasks, the lecturers chose to prioritise practicality over other principles of assessment such as authenticity and washback. The lecturers explained they chose specific assessment task designs like multiple choices and quizzes due to large student numbers, grading time requirements, and the need to give timely feedback, which were found to be some of the massive practical barriers of time constraints and resource limitations (Narreddy *et al*., 2025). This prioritisation of practicality is in line with previous research showing practicality was particularly important for classroom teachers in comparison to other principles of assessment (Brown & Abeywickrama, 2019).

6. Pedagogical Implications

Based on the key findings, some pedagogical implications in an effort to answer the research questions have been made for Vietnamese educators to implement effective FE in tertiary education. First of all, there should be more training to build teachers' capacity in e-assessment design and other related activities (i.e., proctoring, grading, giving feedback). Teachers acknowledge a lack of some competencies and demand pedagogical support (Basilotta-Gómez-Pablos *et al*., 2022), so that they can create high-quality testing and assessment. Secondly, alternatives to e-assessment assignments (e.g., e-portfolios) that focus more on process and higher-order thinking should be utilized accordingly, so that students have to interact more with others and demonstrate other skills (e.g., critical thinking and problem solving). Online tests can be developed by using suitable question formats, cognitive levels, and philosophical approaches, so that task designs are purposeful and authentic in assessing twenty-first century learning (Boitshwarelo et al., 2017). Finally, institutions should provide updated and detailed guidelines on acceptable use of AI for students to complete their assignments and for teachers to do their testing and grading. More instruments should be introduced to assure

student authentication and authorship because cheating behaviours are not as well-controlled as they are in face-to-face proctored assessments (Mellar *et al*., 2018). Importantly, making the washback effects of FE noticeable to the students truly motivates them to study better in online learning environments. A combination of human and automated marking and feedback could be considered for an effective implementation. Online tests can provide "automated rich, descriptive, formative feedback…to scaffold the learning process", enabling students to improve themselves and prepare for summative assessment (Boitshwarelo *et al*., 2017, p. 3). Likewise, peer assessment and peer feedback can complement teacher assessment, particularly in online learning settings (Webb *et al*., 2018), leading to interactive assessment that promotes critical analysis of students' performance (Romeu Fontanillas *et al*., 2016).

7. Conclusions and Recommendations

All in all, this study highlights existing pedagogical challenges faced by Vietnamese educators at the tertiary level when implementing FE. Problems regarding how to maintain validity, reliability, authenticity, practicality, and washback in FE are explicitly addressed in the context of Vietnam's higher education. In response to these pedagogical challenges, suggestions for improvement are offered with a central focus on the implementation of professional support from the institutions, alternatives to conventional e-assessment assignments (e-portfolios under consideration), and guidelines on how to acquire academic integrity as well as how to increase learners' motivation. The generalization of the findings might have some limitations; however, they could be used as a reference for those in similar educational contexts. Further investigations into students' perspectives and the impact of FE on learning outcomes could be conducted in order to gain a comprehensive understanding of FE in Vietnam's tertiary education, contributing to more practical implications to optimize the effectiveness of FE.

Acknowledgements

This research is one of the scientific and technological projects (2024-2025), granted by Vietnam's Ministry of Education and Training under the project code B2024-MBS-02. Valuable support has also been provided by Ho Chi Minh City Open University throughout the implementation of this project. All the contributions are highly appreciated.

References

Aulakh, K., Roul, R. K., & Kaushal, M. (2023). E-learning enhancement through educational data mining with COVID-19 outbreak period in backdrop: A review. *International Journal of Educational Development,* 101, 1-17. https://doi.org/10.1016/j.ijedudev.2023.102814

Bin Mubayrik, H. F. (2020). New trends in formative-summative evaluations for adult education. *SAGE*, 10(3), 1-13. https://doi.org/10.1177/2158244020941006

Basilotta-Gómez-Pablos, V., Matarranz, M., Casado-Aranda, L. A., Otto, A., & García-Peñalvo, F. J. (2022). Teachers' digital competencies in higher education: A systematic literature review. *International Journal of Educational Technology in Higher Education, 19, Article 8.* https://doi.org/10.1186/s41239-021-00312-8

Blanco, M., & Ginovart, M. (2012). On how Moodle quizzes can contribute to the formative e-assessment of first-year engineering students in mathematics courses. *Universities and Knowledge Society Journal*, 9, 354–370. https://doi.org/10.7238/rusc.v9i1.1277

Brinkmann, S. (2018). The interview. In N. K. Denzin & Y. S. Lincoln (Eds.*), The SAGE handbook of qualitative research* (5th ed., pp. 995-1038). SAGE.

Brown, H. D., & Abeywickrama, P. (2019). *Language assessment: Principles and classroom practices* (3rd ed.). Pearson.

Boitshwarelo, B., Reedy, A. K., & Billany, T. (2017). Envisioning the use of online tests in assessing twenty-first century learning: A literature review. *Research and Practice in Technology Enhanced Learning, 12, Article 16.* https://doi.org/10.1186/s41039-017-0055-7

Chen, L.-L., & Chen, C.-M. (2023). Formative

e-assessment design in online learning environments. *International Journal of Education*, *15*(1), 36-49. https://doi.org/10.5296/ije.v15i1.20580

Chen, L., Chen, P., & Lin, Z. (2020). Artificial intelligence in education: A review. *IEEE Access*, 8, 75264–75278. https://doi.org/10.1109/access.2020.2988510

Christiana, M. (2025). Washback in language testing: A literature review of positive and negative influences on teaching and learning. *Jurnal Pendidikan Bahasa*, *15*(3), 397-402. https://doi.org/10.37630/jpb.v15i3.3416

Christodoulou, M. (2025). Grounded theory as a framework for explanatory sequential mixed-method design, an example from educational research. *Quality & Quantity***,** 59, 5557–5573. https://doi.org/10.1007/s11135-025-02214-7

Corwin, B. Z., & Harper, J. (2024). Interviews: Processes, strategies, and reflections. In E. J. Edlund & L. A. Nichols (Eds.), *The Cambridge handbook of research methods and statistics for the social and behavioral sciences*, 2 (pp. 591-615). Cambridge University Press. https://doi.org/10.1017/9781009000796.025

Cotton, D. R. E., Cotton, P. A., & Shipway, J. R. (2024). Chatting and cheating: Ensuring academic integrity in the era of ChatGPT. *Innovations in Education and Teaching International*, *61*(2), 228–239. https://doi.org/10.1080/14703297.2023.2190148

Creswell, J. W., & Creswell, J. D. (2018). *Research design: Qualitative, quantitative, and mixed methods approaches.* SAGE.

Dave, K., & Mitchell, K. (2025). Enhancing flexible assessment through ePortfolios: A scholarly examination. *Journal of University Teaching and Learning Practice*, *22*(3), 1-17. https://doi.org/10.53761/wearsf41

De Villiers, R., Scott-Kennel, J., & Larke, R. (2016). Principles of effective e-assessment: A proposed framework. *Journal of International Business Education*, *11*, 65-92. https://www.neilsonjournals.com/JIBE/abstractjibe11devilliersetal.html

Denzin, N. K. (2014). Writing and/as analysis or performing the world. In U. Flick (Ed.), *The SAGE handbook of qualitative data analysis* (pp. 569-584). SAGE. https://doi.org/10.4135/9781446282243

Fraser, G. (2024). Open-ended survey questions. In E. J. Edlund & L. A. Nichols (Eds.), *The Cambridge handbook of research methods and statistics for the social and behavioral sciences*, 2 (pp. 546-570). Cambridge University Press. https://doi.org/10.1017/9781009000796.025

Gün-tosik, E., Atasoy, B., & Yüksel, A. O. (2023). The role of ePortfolios in formative assessment: A systematic literature review. *Journal of Measurement and Evaluation in Education and Psychology*, *14*, 299-319. https://doi.org/10.21031/epod.1276063

Hagvall Svensson, O., Adawi, T., & Johansson, A. (2022). Authenticity work in higher education learning environments: A double-edged sword? *Higher Education, 84,* 67–84. https://doi.org/10.1007/s10734-021-00753-0

Hai, N. V. (2022). Vietnamese English teachers' views on formative assessment management and suggestions to improve student motivation. *Linguistics and Culture Review*, *6*(1), 363-379. https://doi.org/10.21744/lingcure.v6nS1.2054

Hung, C. T., Wu, S. E., Chen, Y. H., & Chen, P. J. (2024). The evaluation of synchronous and asynchronous online learning: Student experience, learning outcomes, and cognitive load. *BMC Medical Education, 24, 326.* https://doi.org/10.1186/s12909-024-05311-7

Jamalludin, S., & Romli, A. (2023). Research on a learner assessment model in online learning for higher education. In M. Koc, O. T. Ozturk, & M. L. Ciddi (Eds.), *Proceedings of ICRES 2023 International Conference on Research in Education and Science* (pp. 573-587). ISTES Organization. https://files.eric.ed.gov/fulltext/ED654721.pdf

Kearns, L. R. (2012). Student assessment in online learning: Challenges and effective practices. *Journal of Online Learning and Teaching*, *8*(3), 198-208. https://jolt.merlot.org/vol8no3/kearns_0912.pdf

Laato, S., & Murtonen, M. (2020). Improving synchrony in small group asynchronous online discussions. *Advances in Intelligent Systems and Computing, 1161*, 215-224. https://doi.org/10.1007/978-3-030-45697-9_21

Lei, W., & Lei, Z. (2026). Formative assessment literacy: A systematic review. *Language Testing Asia, 16(7)*, 1-18. https://doi.org/10.1186/s40468-025-00418-0

Maleki, A., Khodabakhshzadeh, H., Zeraatpishe, M., & Heidari, F. (2025). Development and validation of the online formative assessment validity scale (OFAVS) in EFL contexts. *Discover Sustainability, 6, Article 471.* https://doi.org/10.1007/s43621-025-01336-0

Ma, M., Ng, T. K. D., Liu, Z., Lin, J., & Wong, K. W. G. (2025). Why don't teachers teach AI ethics? Understanding teachers' beliefs and intentions in Chinese AI curriculum implementation through the theory of planned behaviour. *Computers and Education: Artificial Intelligence*, 9, Article 100518, 1-13. https://doi.org/10.1016/j.caeai.2025.100518

Mellar, H., Peytcheva-Forsyth, R., Kocdar, S., Karadeniz, A., & Yovkova, B. (2018). Addressing cheating in e-assessment using student authentication and authorship checking systems: Teachers' perspectives. *International Journal for Educational Integrity*, *14*(2), 1-21. https://doi.org/10.1007/s40979-018-0025-x

McCallum, S., & Milner, M. M. (2020). The effectiveness of formative assessment: Student views and staff reflections. *Assessment & Evaluation in Higher Education*, *46*(1), 1-16. https://doi.org/10.1080/02602938.2020.1754761

Ministry of Education and Training. (2023). Promulgation of regulations on distance education at the tertiary level, Circular No. 28/2023/TT-BGDĐT [Ban hành quy chế đào tạo từ xa trình độ đại học QĐ 28/2023/TT-BGDĐT]. https://thuvienphapluat.vn/van-ban/Giao-duc/Thong-tu-28-2023-TT-BGDDT-Quy-che-dao-tao-tu-xa-trinh-do-dai-hoc-593889.aspx

Mohammed, R. A., & Mashkhal, A. S. (2023). Evaluating online assessment strategies: A systematic review of reliability and validity in e-learning environments. *North American Academic Research*, *6*(12), 1-18. https://doi.org/10.5281/zenodo.10407361

Mulisa, F. (2022). Sampling techniques involving human subjects: Applications, pitfalls, and suggestions for further studies. *International Journal of Academic Research in Education, 8*(1), 74-83. https://doi.org/10.17985/ijare.1225214

Naeem, M., Ozuem, W., Howell, K., & Ranfagni, S. (2023). A step-by-step process of thematic analysis to develop a conceptual model in qualitative research. *International Journal of Qualitative Methods,* 22, 1-18. https://doi.org/10.1177/16094069231205789

Narreddy, C., Joordens, S., & Prompiengchai, S. (2025). Harnessing large language models for scalable and effective formative assessment in higher education: A review. *Trends in Higher Education*, 4, Article 65, 1-27. https://doi.org/10.3390/higheredu4040065

Nartgün, Z., & Kennedy, E. (2024). Cheating in higher education in the age of artificial intelligence. *International Journal on Lifelong Education and Leadership, 10*(2), 47-54. https://doi.org/10.25233/ijlel.1553831

Ngema, D. (2024). A critical review of the purposes and authenticity of e-assessment policies and practices. In M. L. Mokhele-Makgalwa *et al.* (Eds.), *Proceedings of the 40th AEAA Annual Conference on reimagining educational assessment in the age of multiple dimensions of learning in a global society.* ERRCD Forum. https://doi.org/10.38140/obp2-2024-10

Nhu, H. B., & Tin, D. T. (2019). Impacts of online formative assessment on EFL students' writing achievement. *Ho Chi Minh City Open University Journal of Science - Social Sciences*, *9*(1), 55–69. https://doi.org/10.46223/HCMCOUJS.soci.en.9.1.271

Orsi, M., & Juliano, L. (2021). Impact of formative e-assessment on attendance. *Journal of Higher Education Theory and Practice, 21*(15), 218-224. https://doi.org/10.33423/jhetp.v21i15.4904

Pacheler, N., Daly, C., Mor, Y., & Mellar, H. (2010). Formative e-assessment: Practitioner cases. *Computers & Education*, *54*(3), 715-721. https://doi.org/10.1016/j.compedu.2009.09.032

Padayachee, P., Wagner-Welsh, S., & Johannes, H. (2018). Online assessment in Moodle: A framework for supporting our students. *South African Journal of Higher Education*, *32*(5), 211-235. https://doi.org/10.20853/32-5-2599

Pordanjani, A. Z., & Salehi, K. (2025). Limitations of electronic assessment: A systematic review. *Quanta Research*, *3*(1), 111–130. https://doi.org/10.15157/QR.2025.3.1.111-130

Prendes-Espinosa, M. P., Gutierrez-Porlan, I., & Garcia-Tudela, P. A. (2021). Collaborative work in higher education: Tools and strategies to implement the e-assessment. In R. Babo, N. Dey, & A. S. Ashour (Eds.), *Workgroups e-assessment: Planning, implementing and analysing frameworks*. Springer. https://doi.org/10.1007/978-981-15-9908-8_3

Prendes-Espinosa, P., García-Tudela, P. A., & Gutiérrez-Porlán, I. (2022). Formative e-assessment: A qualitative study based on master's degrees. *International Education Studies, 15*(2), 1-13. https://doi.org/10.5539/ies.v15n2p1

Polaris Market Research Analysis. (2022). *E-learning market share, size, trends, industry analysis report, 2022 – 2030.* https://www.polarismarketresearch.com/industry-analysis/e-learning-market

Romeu Fontanillas, T., Romero Carbonell, M., & Guitert Catasús, M. (2016). E-assessment process: Giving a voice to online learners. *International Journal of Educational Technology in Higher Education*, *13*(20), 1-40. https://doi.org/10.1186/s41239-016-0019-9

Sarıgoz, O. (2023). Teachers' opinions on using web-based e-assessment and evaluation applications in education. *Problems of Education in the 21st Century*, *81*(1), 117-129. https://doi.org/10.33225/pec/23.81.117

Say, R., Visentin, D., Cummings, E., Carr, A., & King, C. (2022). Formative online multiple-choice tests in nurse education: An integrative review. *Nurse Education in Practice, 58, Article 103262*. https://doi.org/10.1016/j.nepr.2021.103262

Singh, V., & Thurman, A. (2019). How many ways can

we define online learning? A systematic literature review of definitions of online learning (1988-2018). *American Journal of Distance Education*, *33*(4), 289–306. https://doi.org/10.1080/08923647.2019.1663082

Trinh, N. M. T., & Trinh, N. B. (2023). Undergraduates' perceptions of online assessment in tertiary education: A case study at Tra Vinh university, Vietnam. *Journal of Contemporary Educational Policies and Practices, 7*(3), 357–364. https://doi.org/10.52296/vje.2023.340

Vo, T. K. A. (2023). E-exams in Vietnam's higher education: Students' computer efficacy and attitudes. *AsiaCALL Online Journal, 14*(2), 62-79. https://doi.org/10.54855/acoj.231425

Wahas, Y. M. A., & Syed, A. J. A. (2024). E-assessment challenges during e-learning in higher education: A case study. *Education and Information Technologies, 29(11)*, 14431-14450. https://doi.org/10.1007/s10639-023-12421-0

Webb, M. E., Prasse, D., Phillips, M., Kadijevich, D. M., Angeli, C., Strijker, A., Carvalho, A. A., Andresen, B., Dobozy, E., & Laugesen, H. (2018). Challenges for IT-enabled formative assessment of complex 21st century skills. *Technology, Knowledge and Learning, 23,* 441–456. https://doi.org/10.1007/s10758-018-9379-7

Yambi, T. (2018). *Assessment and evaluation in education*. University Federal do Rio de Janeiro.

Yan, S., Jiang, L., Zuo, Z., An, L., & Wang, L. (2025). Unravelling and rectifying the free-riding behaviour among students in university flipped classrooms: A uninorm DEMATEL method. *Frontiers in Psychology, 16, Article 1617001. https://doi.org/10.3389/fpsyg.2025.1617001*

Yan, Z., Xiao, Y., Sin, K.-F., Yang, L., & Guo, W.-Y. (2021). Formative assessment practices in special school classrooms with the support of e-books: A case study. *Frontiers in Education, 6, Article 674869*. https://doi.org/10.3389/feduc.2021.674869

Zawacki-Richter, O., Marín, V. I., Bond, M., & Gouverneur, F. (2019). Systematic review of research on artificial intelligence applications in higher education – where are the educators? *International Journal of Educational Technology in Higher Education, 16, Article 39*. https://doi.org/10.1186/s41239-019-0171-0