Exploring the integration and utilisation of generative AI in formative e-assessments: A case study in higher education

Dongpeng Huang, Yixuan Huang, James J. Cummings
Boston University

The integration of generative artificial intelligence (GenAI) into web-based individual formative e-assessments in higher education is a nascent field that warrants further exploration. This study investigated the use of GenAI within an 8-week undergraduate-level research methods course at a university in the United States of America, aiming to understand how students leverage GenAI tools during individual formative e-assessments questions. The research revealed that a significant majority of students initially preferred traditional study resources over GenAI. However, a gradual shift towards more balanced use of both resources was observed, particularly in formative e-assessments involving statistical analysis and calculation questions. In their interactions with GenAI, students primarily used it for multiple-choice and true/false questions, often by directly copying and pasting the question prompt into the GenAI interface. Students were able to discern and accept accurate responses generated by GenAI and reject those that were incorrect or contradicted their existing knowledge. Students' reported primary motivations for turning to GenAI were to seek answers to assessment items as well as to corroborate the accuracy of their own responses. This study contributes to the growing body of literature empirically investigating actual usage behaviours with GenAI tools and the motivation behind these behaviours. We discuss the implications and limitations of these findings.

Implications for practice or policy:

- Educators should develop AI literacy programmes and integrate them into pedagogy strategies.
- Educators and researchers need clear guidelines for ethical AI use in formative e-assessments.
- Educators should encourage students' critical thinking and source evaluation on the information that GenAI provides.

Keywords: individual formative e-assessment, generative artificial intelligence (GenAI), college students, usage behaviour, ChatGPT

Introduction

Generative artificial intelligence (GenAI) consists of machine learning models that are capable of creating new content based on massive training data sets (BaiDoo-Anu & Owusu Ansah, 2023; Jovanovic & Campbell, 2022). In recent years, GenAI has taken a significant leap into higher education, offering innovative tools that could revolutionise the learning process and present both opportunities and challenges for educators and students alike. GenAI has been lauded for their potential to support personalised learning paths (Alasadi & Baiz, 2023), promote accessibility (BaiDoo-Anu & Owusu Ansah, 2023) and provide instant academic assistance (Luckin et al., 2016). Students utilise GenAI to overcome language barriers (Pack & Maloney, 2023), generate adaptive learning materials (Alasadi & Baiz, 2023) and complete assessment questions on various study contents (Voss et al., 2023). The motivations behind students' use of GenAI for study are diverse and contextual, ranging from seeking help with comprehension to meeting needs for efficient information retrieval (Hmoud et al., 2024; Hou et al., 2024). For instance, students employ GenAI to address open-ended essay-writing tasks, enjoying benefits that include pleasure and satisfaction from task completion, time and effort conservation, enhancement of self-assessment capabilities, perceived usefulness and the opportunity to engage in conversations and receive feedback (Hmoud et al., 2024). When engaging in computer programming tasks, students turn to GenAI to learn new concepts, write code, debug and develop test cases (Hou et al., 2024).

Studies have yielded mixed findings on the impact of GenAI on students' learning outcomes. Some research has reported significant positive effects on motivation enhancement (Yin et al., 2021), self-efficacy (Yilmaz & Yilmaz, 2023), knowledge acquisition (Lee et al., 2022) and skill development in areas such as listening (Chien et al., 2022), speaking (C.-J. Lin & Mubarok, 2021) and writing (M. P.-C. Lin & Chang, 2020). Conversely, other studies have reported that GenAI did not significantly impact the student learning process in aspects like reading engagement (Liu et al., 2022) or motivation and self-efficacy (Kumar, 2021), or that its effectiveness depended on moderators such as students' initial perceived choice (Yin et al., 2021) and duration of use (Wu & Yu, 2024).

Given these varied findings, researchers have recently considered the relative promise of GenAI for augmenting traditional teaching methods (Alasadi & Baiz, 2023) and complementing traditional learning resources (Cooper, 2023). For instance, Alasadi and Baiz suggested that GenAI could compensate for the conventional limitation of not having multiple teachers per classroom by serving in a cost-effective teaching assistant role. Moreover, Cooper posited that GenAI could assist students in reviewing new concepts before engaging with traditional learning resources, such as textbooks and lectures. However, a recent study concluded that GenAI cannot entirely replace traditional learning sources (Hou et al., 2024). Hou et al. empirically investigated students' help-seeking behaviour in computing tasks with GenAI and identified barriers – such as formulating effective help requests, and a lack of usage experience and trust in the models – that hinder GenAI from being an effective learning resource. With Hou et al.'s study as a rare exception, there remains a lack of empirical research and contextual exploration into students' preferences for GenAI versus traditional learning resources and the decision-making processes behind their choices. Consequently, there is a call for more rigorous integration and evaluations to ascertain the impact of GenAI in more specific educational contexts (Darwin et al., 2024). This study contributes to the body of knowledge by focusing on student's GenAI usage in individual formative e-assessment.

Individual formative e-assessment and learning

Individual formative e-assessment refers to the evaluation of a student's learning performance through tasks such as online quizzes consisting of multiple-choice or true-false questions, utilising web tools throughout the learning process (Cohen & Sasson, 2016). Such assessments involve generating evidence about a learner's progress and enabling actions that have formative effects on their educational journey (Daly et al., 2010). Formative e-assessment is a personalised approach to gauge a student's progress, understanding, strengths and areas that need improvement within an academic setting or specific course (Stödberg, 2012), serving as a crucial feedback mechanism for both students and instructors in college education. It is argued that well-designed formative e-assessments can motivate students to engage more deeply with the course content (Walker et al., 2008) and to allocate their study efforts more effectively (McCallum & Milner, 2021). Studies have found that immediate and constructive feedback from formative e-assessments allows students to identify their weaknesses and work on them, leading to improved performance in subsequent formative e-assessments (Hettiarachchi et al., 2015; Jiao, 2015). Formative e-assessment also encourages students to self-regulate their learning by planning (García-Jiménez, 2015), monitoring and developing learning strategies (Mohamed, 2020). For instructors, studies indicated that formative e-assessments provide insight into how well the students are grasping the course materials, enabling them to adjust teaching strategies or materials to reach their teaching goals (Al-Hattami, 2020; Bahati et al., 2019). Educational practices often adopt an open-book and open-web approach for e-assessment because it is closer to the real-life learning environment (Marimuthu & Ramraj, 2019). Generally speaking, open-book and open-web describes a situation where students are permitted to use any required resource materials during an exam (Joshi et al., 2020). To this end, this study defines any learning resource that students could refer to prior to the emergence of GenAI as a traditional learning resource, in contrast to the new learning resource of GenAI.

Research on using GenAI tools in individual e-assessment has focusds primarily on encouraging instructors to take advantage of GenAI on quiz questions design (Bozkurt et al., 2023; Cooper, 2023) or generating explanations for quiz answers (McDonald et al., 2024). Its usage in formative e-assessments raises important considerations on academic integrity and the validity of e-assessments practices. For example,

George (2023) posited that GenAI could potentially support fair and personalised formative e-assessments, but educators and learners should be cautious against the risks of over-reliance on GenAI, which might undermine critical thinking skills and encourage academic dishonesty. Moving towards more authentic formative e-assessments, educators need clear guidelines, increased labor involved and the potential need for changes in class sizes, teaching loads or grading support (Hodges & Ocak, 2023).

Although research has shown positive student feedback on the enhancement of learning experiences through GenAI tools (Chan & Hu, 2023), there is limited knowledge regarding students' actual usage behaviours. Understanding the nature of questions directed to GenAI, the prompts employed by students and the accuracy of the responses received is crucial in evaluating GenAI's functional impact on assessment outcomes and learning achievements. This study is among the first to explore the real-time behaviours and decision-making processes of students engaging with GenAI during individual formative e-assessments. The outcomes of this research could lead to a more profound understanding of college students' actual usage behaviours with GenAI and the potential benefits and challenges of incorporating GenAI into higher education – a practice that is still emerging but rapidly gaining importance. Additionally, these findings could help inform policies and guide future pedagogical strategies to ensure the integration of AI in educational settings is both efficient and ethically sound. Consequently, we posed the following research questions (RQs) to explore how students utilise GenAI in individual information-recall formative e-assessments:

- RQ1: Do students prefer to use traditional learning resources or GenAI for assistance with individual formative e-assessment questions?
- RQ2: What motivates students to use GenAI in their formative e-assessments?
- RQ3: What are the general usage patterns of students using GenAI to assist them with individual formative e-assessment questions?
- RQ4: Why do students accept or reject answers provided by GenAI?

Method

Design

This exploratory study took place during a 14-week undergraduate communication research methods course at a large north-eastern university in the United States of America, involving 8 weeks of data collection. The course aimed to familiarise students with diverse research methodologies such as surveys, content analyses and experiments, equipping them to discern and apply appropriate research methods. The university's Institutional Review Board approved the study, which consisted of four formative e-assessments where students used ChatGPT alongside traditional learning resources.

ChatGPT was chosen for its widespread availability and popularity among students at the time of the study in 2023 and high-quality, context-aware responses, which simplify user interactions by allowing challenges to incorrect assumptions and handling inappropriate requests (Haleem et al., 2022; Lock, 2022). These elements provide an ease of use that has positioned ChatGPT as a leading tool (Westfall, 2023). Moreover, using ChatGPT standardised the GenAI tool across the study, limiting variables like differing GenAI capabilities or user interfaces, thereby enhancing the study's internal validity. Further, technical comparisons suggest ChatGPT's features are on par with major GenAI services from companies like Microsoft and Google, suggesting that findings from this study could apply broadly to similar GenAI tools (Marr, 2024; Raffo, 2024).

To ensure equal opportunity to engage with GenAI without financial or other access barriers, all students were provided with a subscription to the paid version of ChatGPT 4.0. To let the students know that the use of GenAI in the course was permissible, the policy was explicitly specified in the syllabus as follows:

> This course examines developing technologies including generative AI and relies on students to be responsible in using technology to develop their research skillsets in and out of the classroom. However, when using generative AI in assessments, assignments, and group work, it is essential to disclose the prompt/s you used and provide appropriate acknowledgment of the AI tool/s using APA Style. This specific approval is given to allow you to become familiar with AI tools, to increase your AI literacy, and to understand how to use AI in appropriate ways within an academic context. Details will be given in class. Submitting any AI-generated content as an originally-created essay, paper, project, or work is strictly prohibited.

All e-assessments were auto-graded in real time immediately upon submission. Following the completion of each e-assessment, students filled out an online survey questionnaire to report their ChatGPT usage behaviours, including the accuracy of their assessment answers.

A total of four formative e-assessments were conducted every 1–2 weeks throughout the semester via the Blackboard online learning management platform. The e-assessments consisted of in-class, time-restricted questions aimed at evaluating recall of the course material. The questions were adopted from a previous instance of the same course. They were designed to gauge conceptual and practical comprehension of different research methodologies and statistics taught in lectures. Each e-assessment contained between 20 and 25 questions, encompassing four question types that are typically used to assess students' learning outcomes: multiple-choice, true/false, fill-in-the-blank and matching. The instrument used for assessment in this course was taken directly from a prior version of the course. Table 1 presents detailed information on the content covered in each e-assessment and the distribution of question types.

Table 1
E-assessment question types

	Content overview	**MC**	**T/F**	**Fill-in-the-blank**	**Matching**
Assessment 1 (21 questions)	Information search techniques. Identification of disinformation and misinformation. Basics of research design. Communication research models. Measurement tools: index, scale and typology. Measurement principles: validity, reliability.	10	7	4	0
Assessment 2 (23 questions)	Survey methodology. Questionnaire design and testing. Sampling techniques.	12	9	1	1
Assessment 3 (22 questions)	Basic statistics. Hypothesis testing. Research ethics and Institute Review Boards. Longitudinal study; data collection and management; existing data banks & secondary analysis. Experiment methods and design. Qualitative methods.	10	7	5	0
Assessment 4 (24 questions)	Everything above plus content analysis method.	8	10	5	1

Note. MC = multiple-choice; T/F = true/false.

Participants

The participants comprised 25 undergraduate students with the following demographic breakdown: gender: 80% female, 16% male, 4% identified with a self-described gender identity; age: ranged from 20 to 23 years (M = 20.44, SD = 0.77); ethnicity: 80% Asian or Pacific Islander, 12% Caucasian, 8% Hispanic. Of these students, 88% (n = 22) reported having prior experience with GenAI. Within the opening weeks of the class, all students completed a ChatGPT tutorial, resulting in all students having direct experience with and knowledge of its basic functionality prior to the beginning of the study data collection period.

Procedure

Students were required to complete the formative e-assessments individually within a 45-minute time frame. They were permitted to utilise textbooks, notes, search engines, ChatGPT and calculators as resources. The relevant instruction in the syllabus stated:

> There will be 4 assessments posted on Blackboard that aim to ensure that you follow and thoroughly comprehend the subject of the specific areas. Each of your assessments will be composed of 20-25 questions and conducted in class. All the assessment results will be included for your final grade at the end of the semester. You can have access to ChatGPT, books, notes, and calculators when taking the assessment; and you will need to finish it within 45 minutes.

Upon completing the formative e-assessments, students were asked to complete an online survey questionnaire to report their usage of ChatGPT, including the specific assessment questions addressed, the prompts posed, the reasons for adopting or not adopting ChatGPT's answers (including copying partially or in full the AI-generated answers to fill in question blanks, as well as mapping ChatGPT's generated answers to options provided when selecting an answer to a multiple-choice question). Additionally, participants reported the score they then received on each of the auto-graded assessments.

Results

Rate of ChatGPT usage

Excluding a single fill-in-the-blank question in Assessment 2, where no participant utilised ChatGPT, each question type across all formative e-assessments experienced at least some degree of ChatGPT usage by participants. Given the unequal number of questions in each type across formative e-assessments and the observation that ChatGPT was employed by students for every question type, for each assessment we calculated the percentage of questions of each type that were answered with the use of ChatGPT. As presented in Table 2, multiple-choice and true/false questions exhibited the highest rate of ChatGPT usage, indicating these are the question types most commonly addressed with the assistance of ChatGPT.

Table 2
Rate of ChatGPT usage per question types across four formative e-assessments

	Multiple-choice	True/False	Fill-in-the-blank	Matching
Assessment 1	44.00%	35.43%	27.00%	0.00%
Assessment 2	34.33%	36.44%	0.00%	40.00%
Assessment 3	45.20%	30.86%	27.20%	0.00%
Assessment 4	42.00%	30.40%	33.60%	36.00%

To formally test the assessment number and the type of questions contained in each assessment — as well as the potential interaction of these factors — on students' ChatGPT usage, a mixed effects ordinal logistic regression was conducted to investigate how the likelihood of ChatGPT usage in a given response (N = 2,250) varied by assessment and the type of question. Assessment number and question type were included as fixed effects factors while additional variance between participants completing multiple

responses was accounted for as a random effect. The reference level for the assessment number is Assessment 1 (basic concepts of research methodology). The reference level for question type is multiple-choice question. The model is presented in full in Table 3, with the effect of predictors on the likelihood to use or not use ChatGPT presented as odds ratios. The results showed that Assessment 2 (survey methodology) has significantly lower odds of including the use of ChatGPT compared to Assessment 1 (*OR* = 0.67, *p* = .001). Additionally, compared to multiple-choice questions, the odds of using ChatGPT were significantly lower for true/false questions (*OR* = 0.67, *p* < .001) and for fill-in-the-blank questions (*OR* = 0.46, *p* < .001).

Table 3
The likelihood of ChatGPT usage across conditions

Parameter	Odds ratio	SE	CI	*p*
Assessment 2	0.67	0.16	0.50–0.92	0.01*
Assessment 3	0.86	0.15	0.64–1.17	0.34
Assessment 4	0.84	0.17	0.61–1.17	0.31
True/False	0.67	0.11	0.54–0.83	0.00*
Fill-in-the-blank	0.46	0.15	0.34–0.61	0.00*
Matching	0.93	0.33	0.49–1.79	0.83

*$p \leq .05$.
Note. Student's ID was included as a random effect.

Traditional resources versus ChatGPT

To answer RQ1, we compared students' reliance on ChatGPT versus traditional study resources across the four formative e-assessments, which revealed a fluctuating trend (see Figure 1). In the first assessment (about basic concepts of research and measurement), a substantial majority of students (60.10%) favoured traditional resources over ChatGPT (39.90%). Assessment 2 (about survey methodology) also demonstrated an overwhelming preference for traditional methods (68.87%) and less reliance on ChatGPT (30.43%). However, Assessment 3 (about longitudinal study, experiment and qualitative methods) showed a more balanced distribution, with 51.82% of students utilising traditional resources and 47.09% opting for ChatGPT. The disparity then widened again in Assessment 4 (everything covered by the first three assessments plus content analysis method), where 56.00% of students reverted to using traditional resources, compared to a smaller proportion (43.00%) utilising ChatGPT. Notably, the percentage of students reporting equal use of both ChatGPT and traditional resources was negligible in Assessments 2, 3 and 4 (0.70%, 1.09% and 1.00%, respectively), suggesting a clear preference for one method over the other in these instances.

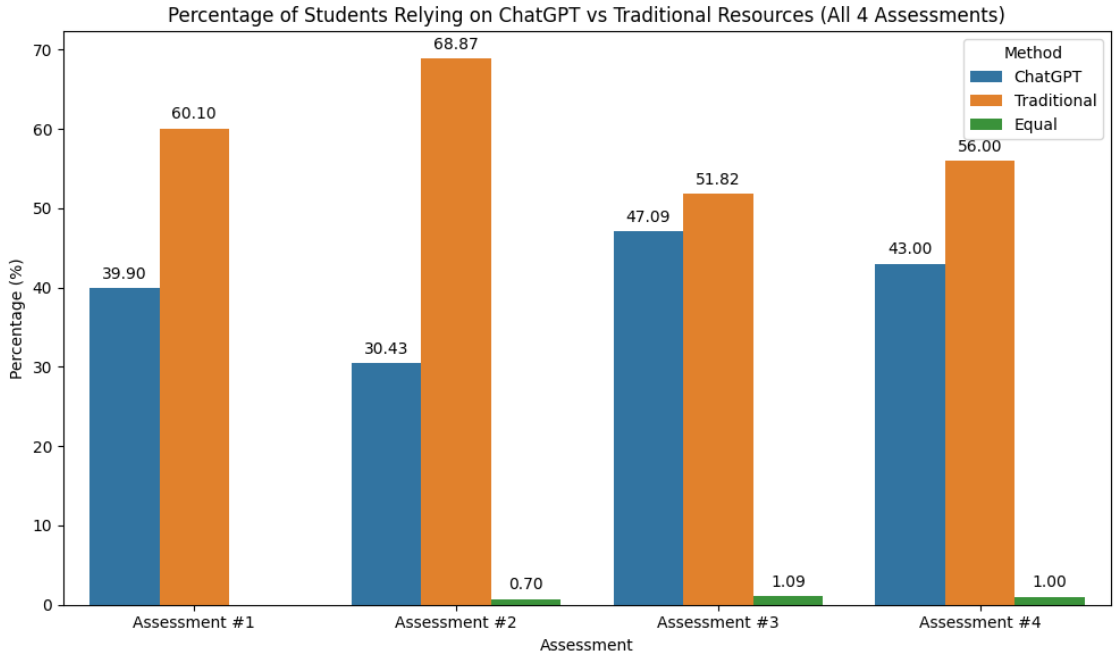

Figure 1. Percentage of students using traditional resources or ChatGPT across four assessments

Motivations for ChatGPT use

To answer RQ2, we conducted a content analysis using the conventional approach outlined by Hsieh and Shannon (2005) to further identify students' motivations for utilising ChatGPT. According to Hsieh and Shannon, this approach is designed to gain direct insights from students' open-ended question answers without imposing preconceived categories. Researchers read the text word by word, highlighting essential concepts that form the basis for codes. During this process, the researcher documents initial impressions and interpretations. As codes are identified, they are grouped into categories based on their relationships and shared themes. One coder coded for this section and came up with a coding scheme containing four distinct motivations for using ChatGPT. Another second rater then coded a subset of the total data (140 out of 235 participant responses) using the same scheme. The second rater coded more than 10% of the entire data set, which is considered acceptable practice (O'Connor & Joffe, 2020). Intercoder reliability was assessed using Cohen's kappa (κ). The agreement between the two coders was substantial, $\kappa = 0.872$, $p < .001$, indicating a high level of consistency in the coding process. The four reasons for rejecting ChatGPT answers are explained below.

Motivation 1 – Gaps in knowledge. A significant number of students stated that they resorted to ChatGPT either because they were uncertain of their answers or were not familiar with the concepts presented in the quiz questions. Students gave the following reasons for using ChatGPT: "I didn't know", "I had no idea" and "I don't know the answer".

Motivation 2 – Need for clarification. Students suggested using ChatGPT to search for definitions of concepts being asked in the assessment question prompt or for more comprehensive explanations of specific terms or research concepts. For example, students indicated, "I was unsure of the correct answer and wanted to look up definitions of each term and decide on an answer on my own", and "I wanted a definition for paradigm that was more detailed than what I had written in my notes".

Motivation 3 – Verification purposes. Students mentioned using ChatGPT to validate their answers or reasoning. For example, students noted, "I wasn't sure about the answer, wanted to verify it" and "I'm not sure about the first answer. Even though I know about the nomothetic approach, I'm not sure if Choice A would be a good example".

Motivation 4 – Convenience of obtaining quick responses. Students also suggested that they used ChatGPT because of its capability to provide prompt and detailed responses. For example, they indicated, "It is much easier and quicker", "Because I am tired and I want to hear from ChatGPT about the answer easily" and "T/F question is easier to answer for (Chat)GPT".

Based on the topics identified in examining students' stated motivations for ChatGPT use in completing Assessment 1, as well as the potential for respondent fatigue in light of the original question format, we then revised the format of the question for Assessments 2 to 4 to multiple-choice. Students were asked, "Why did you decide to use ChatGPT for this question?"; possible response options included, "To find the answers", "To verify answers I already had", "To find examples that aid in better understanding the concepts in the question prompt" and "Other reason, please specify". Distribution of responses for Assessments 2 to 4 is illustrated in Table 4. Notably, only six students selected the "other" category, all with respect to completing Assessment 2. Three of them explicitly mentioned the following reasons: "I could not find information in my notes", "I didn't understand what the question meant so ChatGPT helped fill in my understanding" and "I didn't understand one of the answers". The other three answers were not meaningful, indicating that they mixed this question up with other questions (e.g., "Oops I mixed this one up with the last question") or did not use ChatGPT but chose the wrong answer in the previous step (e.g., "Ah sorry! I guess I didn't use ChatGPT for this – I thought I did").

Table 4
Students' motivations for using ChatGPT

	To find answers (counts)	**%**	**To verify answers (counts)**	**%**	**To find examples (counts)**	**%**	**Other reasons (counts)**	**%**
Assessment 2 (*n* = 244)	132	54.1%	94	38.52%	12	4.92%	6	2.46%
Assessment 3 (*n* = 182)	116	63.74%	56	30.77%	10	5.49%	0	0.00%
Assessment 4 (*n* = 178)	127	71.35%	39	21.91%	12	6.74%	0	0.00%

ChatGPT usage patterns

To answer RQ3, after students identified the questions for which they used ChatGPT for assistance, they were systematically prompted to provide further details about their usage (see Figure 2).

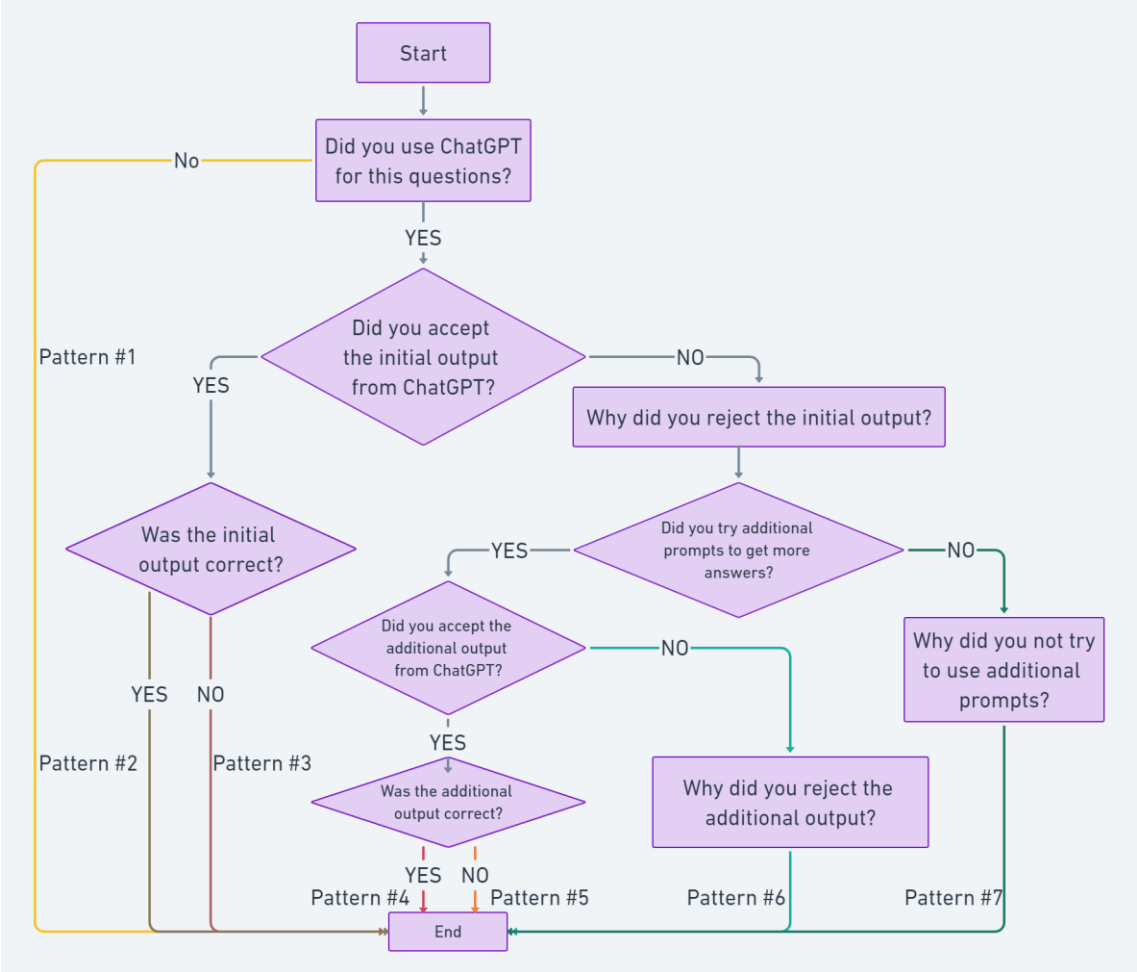

Figure 2. Flowchart of questions asked to assess students' ChatGPT usage practices

The results revealed seven distinct patterns of ChatGPT usage for assistance with formative e-assessment questions:

(1) The student did not use ChatGPT.
(2) The student engaged with ChatGPT on the first attempt, chose to use the output provided and the answer was correct.
(3) The student utilised ChatGPT on the first attempt, accepted the provided output, but the output was incorrect.
(4) The student initially used ChatGPT, did not use the suggested output, but on a subsequent attempt, decided to use the ChatGPT output, which was correct.
(5) The student initially engaged with ChatGPT, did not adopt the suggested output, made a second attempt and adopted the ChatGPT output, but it was incorrect.
(6) The student initially used ChatGPT and did not adopt the suggested output, followed by a second use of ChatGPT where the suggested output was again not adopted.
(7) The student made initial use of ChatGPT, chose not to adopt the suggested output and did not attempt to use ChatGPT a second time.

The frequency of each usage pattern across each e-assessment is detailed in Table 5.

Table 5
Frequency of each student usage pattern across four formative e-assessments

Usage patterns	**Assessment 1 (*n* = 525)**		**Assessment 2 (*n* = 575)**		**Assessment 3 (*n* = 550)**		**Assessment 4 (*n* = 600)**	
	Count	%	Count	%	Count	%	Count	%
#1. Not Used	285	54.29%	367	63.82%	337	61.27%	349	58.17%
#2. Initial Use, Adopted, Correct	177	33.71%	166	28.87%	174	31.64%	168	28.00%
#3. Initial Use, Adopted, Incorrect	18	3.43%	29	5.04%	22	4.00%	40	6.67%
#4. Initial Use, Not Adopted, Second Attempt Adopted, Correct	11	2.10%	3	0.52%	5	0.91%	5	0.83%
#5. Initial Use, Not Adopted, Second Attempt Adopted, Incorrect	0	0.00%	0	0.00%	0	0.00%	0	0.00%
#6. Initial Use, Not Adopted, Second Attempt, Not Adopted	3	0.57%	1	0.17%	0	0.00%	1	0.17%
#7. Initial Use, Not Adopted, No Second Attempt	31	5.90%	9	1.57%	12	2.18%	6	1.00%

Note. Within each column, the value for *n* is the sum of ChatGPT use pattern counts across all seven patterns for that assessment. It equals the number of questions in the assessment multiplied by the number of students taking that assessment.

The results revealed that over half of the students did not utilise ChatGPT for assistance with their e-assessment questions. Among those who did use ChatGPT, a significant proportion adopted the AI's answers, which turned out to be correct. A small minority of students chose not to adopt ChatGPT's answers after either the first or second attempt. Notably, there were no instances of students reporting the adoption of incorrect answers provided by ChatGPT.

The rate of adopting ChatGPT's answers after the first attempt was notably high across all four formative e-assessments: 88.94% for Assessment 1, 85.64% for Assessment 2, 86.07% for Assessment 3 and an impressive 98.25% for Assessment 4. For individuals who chose not to adopt ChatGPT's answers after the first attempt, the likelihood of attempting a second time was generally low, except for Assessment 4, where all three students who did not adopt the initial answers proceeded to make a second attempt.

Specifically, the percentages of students who tried a second time were 4.55% for Assessment 1, 3.57% for Assessment 2, 10.71% for Assessment 3 and 100% for Assessment 4.

Strategies for generating queries

When utilising ChatGPT to assist with individual formative e-assessments, students employed various strategies for generating their initial queries. The majority of students either directly copied and pasted the question text or took a screenshot of the assessment question and submitted it as a ChatGPT query, a practice observed across different types of questions. Specifically, for multiple-choice questions, strategies ranged from copying the complete question with choices included to copying merely the question text, excluding the choices. Additionally, some students sought clarification on the accuracy or interpretation of particular options. Students also copied the true/false questions prompt with or without choices. However, they also tried to rephrase the original question prompt into a new query or probing specific concepts mentioned in the question prompt. For fill-in-the-blank questions, approaches diverged between copying the full question and formulating queries that incorporated a "what" in place of the blank. In the context of matching questions, students either replicated the entire question, including all potential matches, or inquired about the correctness of a presumed pairing. Illustrative examples of these strategies are presented in Table 6.

Table 6
Examples of students' query generation strategies

Assessment question type	**Example assessment question**	**Initial query strategy**	**Example query**
Multiple-choice	What kind of journal is Variety? a) refereed journal b) trade journal c) popular magazine d) tabloid	Copy the entire question with choices included.	What kind of journal is Variety? a) refereed journal b) trade journal c) popular magazine d) tabloid
		Copy the entire question with choices excluded.	What is a refereed journal?
		Ask the accuracy or interpretation of particular options.	What is a refereed journal?
True/False	Nexis Uni provides a database of recorded video of newscasts on network and cable television channels. True False	Copy the entire question with choices included.	Nexis Uni provides a database of recorded video of newscasts on network and cable television channels. True or False?
		Copy the entire question with choices excluded.	Nexis Uni provides a database of recorded video of newscasts on network and cable television channels.
		Paraphrase the original question into a new query.	Does Nexus uni have databases of recorded newscast video?
		Asked specific concept mentioned in the questions prompt.	What is Nexis Uni?

Assessment question type	Example assessment question	Initial query strategy	Example query
Fill-in-the-blank	_____ analysis is a research alternative that utilises available data stored in archives. It is inexpensive and allows researchers to focus more on analysis rather than data collection. What word should be placed in the blank?	Copy the entire question.	_____ analysis is a research alternative that utilises available data stored in archives. It is inexpensive and allows researchers to focus more on analysis rather than data collection.
		Insert "what" in place of the blank.	What analysis is a research alternative that utilises available data stored in archives?
Matching	Match items on both sides • party affiliation • temperature • annual income • socioeconomic status (low - middle - high) A. nominal level B. ordinal level C. interval level D. ratio level	Copy the entire questions including all matching choices	Match items between the lists • party affiliation • temperature • annual income • socioeconomic status (low - middle - high) A. nominal level B. ordinal level C. interval level D. ratio level
		Asked presumed pairings.	Is party affiliation nominal?

As previously noted, the majority of students refrained from posing secondary questions when utilising ChatGPT for assistance with their formative e-assessments (see Table 5). However, among the relatively small number of students who did engage in follow-up queries, the predominant requests were for further explanation or a repetition of the initial query. Notably, some students copied the entire question, including choices, in their initial query and subsequently inquired about a specific choice in the follow-up question for verification purposes.

Reasons for rejecting ChatGPT's answers

To answer RQ4, we conducted a content analysis using the conventional approach outlined by Hsieh and Shannon (2005) to further identify the primary reasons why students chose to reject ChatGPT's answers for assessment questions. One coder coded for this section and came up with a coding scheme comprised of five distinct reasons for rejecting ChatGPT answers. A second rater then coded a subset of the total data (69 out of 87 participant responses) using the same scheme. Intercoder reliability was assessed using Cohen's kappa (κ). The agreement between the two coders was substantial, $\kappa = 0.854$, $p < .001$, indicating a high level of consistency in the coding process. The five reasons for rejecting ChatGPT answers are explained below:

Reason 1 – ChatGPT output deemed insufficient. Students reported that ChatGPT's responses often did not meet their expectations or address the specifics of their queries. They suggested that ChatGPT tended to provide broad answers that lacked the detailed information or explanations they were seeking. Furthermore, the responses generated by ChatGPT were sometimes unrelated to the context of the question, and ChatGPT failed to select a single best option when the prompt implicitly required it, leading

to a presentation of all options as correct. For instance, one student mentioned, “ChatGPT gave an answer that doesn’t relate to what I’m asking”, and another observed, “Since I forgot to specify ‘Choose the right answer,’ ChatGPT did not give me one option within the four candidates. Instead, ChatGPT indicated all four are correct. So, I referred to my notes and identified the correct answer”.

Reason 2 – ChatGPT output incomplete. Instances were reported where ChatGPT was unable to generate a response, either due to insufficient information or limitations inherent to its programming. Examples provided by students include: “Because it said it couldn’t give an answer due to not enough information,” and “ChatGPT didn’t provide an answer, so I decided to guess on my own.”

Reason 3 – ChatGPT output deemed incorrect. Students perceived ChatGPT’s answers as incorrect or incongruent with their existing knowledge base, traditional resources (e.g., classroom instruction, textbook content) or alternative sources (e.g., Google or Quizlet). For instance, responses included, “Because it didn’t seem correct; it felt somewhat off based on our current learning” and “Because my notes provided a different answer, and I trust my notes completely”.

Reason 4 – ChatGPT output did not clearly map onto assessment response options. ChatGPT may provide an answer that does not match any of the options given in a multiple-choice or similar format question. Examples include, “It wasn’t one of the answer choices”, “ChatGPT returned an answer that was not one of the options, so I decided to answer it on my own instead of asking the question again”,

Reason 5 – ChatGPT used only for corroboration. Students rejected ChatGPT’s answer because their searching goal is only to verify the correctness of their own answers or gain additional perspectives on a topic, for example, “I wanted to double check to see if it had other ideas” and “I had another answer and wanted to compare mine with it”.

Discussion

Recent scholarship has suggested that integrating GenAI into college classrooms holds the potential to augment traditional learning resources and enhance students’ learning experiences. However, there is a lack of empirical studies explaining students’ actual usage behaviour within a real classroom setting, especially in formative e-assessment. This study sought to address this gap by conducting a longitudinal exploratory study in an undergraduate college course to examine students’ actual usage behaviours, their preference for traditional learning resources versus ChatGPT, motivations for usage, their reasons for rejecting ChatGPT’s outputs and their strategies for generating queries. Detailed explanations are provided below.

ChatGPT usage rates are higher on multiple-choice and true/false questions

The study results reveal a notable trend where multiple-choice and true/false questions exhibit the highest rate of ChatGPT usage among students. This trend suggests a preference for or a perceived utility in leveraging AI tools for questions that require selection from given alternatives, as opposed to those necessitating the generation of unique answers or associations. It is possible that multiple-choice and true/false questions, by their nature, offer a finite set of possible answers, which could align well with the capabilities of GenAI tools like ChatGPT. It contrasts with fill-in-the-blank and matching questions, which may require more contextual understanding that students might prefer to address through traditional study methods or other resources. Traditional learning materials such as class notes and personalised study materials convey the perspectives of instructors or the students themselves. As these materials are tailored to meet specific course objectives or individual learning styles, they may influence students’ preference for them to address class-specific contextual assessment questions. Additionally, this pattern of ChatGPT usage could reflect students’ strategic approach to optimising their performance on formative e-assessments. Given the pressure to achieve high marks, students may prioritise the use of ChatGPT for question types where the tool’s assistance is perceived to be most effective or where the question format allows for rapid confirmation of answers.

Traditional resources are preferred over ChatGPT

Addressing RQ1, the results show evolving student preferences between ChatGPT and traditional resources across the four formative e-assessments. In Assessment 1, most students preferred traditional resources over ChatGPT, suggesting caution in integrating AI into study practices. This preference is likely due to the factual and conceptual nature of the questions, which led students to trust textbooks and lecture notes more. This trend persisted in Assessment 2, with a slight increase in ChatGPT usage as students became more familiar with the tool. In Assessment 3, a near-equitable preference emerged, possibly due to the introduction of statistical analysis or calculation questions, encouraging students to use ChatGPT for explanations and assistance. However, in Assessment 4, the preference shifted back to traditional resources, likely due to fewer statistical questions. The negligible percentage of students using both resources equally highlights a tendency to prioritise one method over a hybrid approach.

ChatGPT use is motivated by knowledge gap and verification purposes

Regarding RQ2, the findings on students' motivations for utilising ChatGPT in their formative e-assessments shed light on the role of GenAI in addressing knowledge gaps and enhancing learning efficiency. Across all four formative e-assessments, the most commonly mentioned motivation is to find the answers to the assessment questions, suggesting that students use ChatGPT as an immediate resource to bridge the knowledge gaps in their understanding or familiarity with the course material. Students' reliance on ChatGPT for definitions and detailed explanations of course concepts highlights its role not just as a tool for answer retrieval but also as an educational aid that facilitates deeper engagement with the material. Moreover, the use of ChatGPT for double-checking and verification purposes reflects a strategic approach to learning, where students leverage GenAI to reinforce their understanding and confidence in their responses. This aspect of ChatGPT's utility underscores its value in fostering a reflective learning environment, encouraging students to critically assess their own knowledge and assumptions. The mention of ChatGPT's ability to provide quick and thorough answers also highlights the efficiency of AI tools in education. Although students expressed concerns about the accuracy of the information provided by ChatGPT, the capability to access information rapidly still renders it an attractive option for those aiming to optimise their study time and resource usage.

When used, ChatGPT results in high rate of adopting correct – and not incorrect – answers

With respect to RQ3, the study identified seven distinct usage patterns among students using ChatGPT for formative e-assessments. A significant finding is the high rate of adoption of ChatGPT's answers, with most being correct across all assessments. Out of 794 instances of students accepting the generated answer from the initial inquiry, 109 cases (13.7%) involved incorrect answers. No students reported accepting incorrect answers on their second attempt, indicating either student proficiency in evaluating ChatGPT's responses or the tool's high reliability within the assessed subjects.

The adoption rate, especially the 98.25% in Assessment 4, suggests growing student confidence in ChatGPT's accuracy. This trend may reflect GenAI advancements or increased student familiarity and trust in the technology. This observation holds promise for integrating GenAI tools in education, suggesting that judicious use may enhance learning without misleading students. Among 2,250 total question responses, 912 (40.5%) involved using ChatGPT, suggesting that students often felt confident completing questions without GenAI assistance. In 87 cases (9.54%) where the initial output was not accepted, time constraints of the assessments may have limited repeated use.

Strategies for generating queries include direct copying, rephrasing and follow-up questions

Students commonly employed a straightforward method to seek assistance from ChatGPT, often directly copying and pasting or taking screenshots of assessment questions to input into ChatGPT. This approach suggests that students are mainly in search of direct answers or confirmations to the questions they face,

taking advantage of ChatGPT's ability to quickly process and respond to precise queries. Nevertheless, the variability in students' approaches to different types of questions, such as whether to include or exclude multiple-choice options or how they rephrase true/false and fill-in-the-blank questions, indicates a more refined level of interaction with specific types of questions. Furthermore, the strategic use of follow-up questions that probe into particular choices after an initial query demonstrates that students are not just seeking answers. They are also intent on understanding the logic behind specific options and delving into the concepts underlying the questions. This indicates that students are using ChatGPT not just as an answer machine, but as a tool to deepen their understanding and reinforce their learning.

Rejecting ChatGPT's answers is due to irrelevant, incomplete or incorrect answers

Speaking to RQ4, the findings show that students reject ChatGPT's answers mainly due to the perceived quality and their usage purpose. Rejections often stem from unsatisfactory answers, whether irrelevant, false or lacking. Students' frustration with broad or contextually irrelevant answers underscores the need for more sophisticated GenAI tools that can better understand and adapt to specific academic queries. This highlights the necessity for GenAI development in tailoring response lengths and ensuring clarity based on the query context. Improving GenAI's knowledge base and its ability to navigate constraints could help in setting realistic expectations for students regarding GenAI's capabilities and limitations.

Students use ChatGPT for verifying answers or gaining additional perspectives but reject responses inconsistent with traditional knowledge sources like textbooks or class notes. This suggests that while ChatGPT is trusted for corroboration, it is distrusted when conflicting with established sources. This possibly indicates that students possess necessary information literacy, understanding GenAI's limitations and using it as one of many learning tools. However, this situation also presents an opportunity for educators to guide students on effectively using GenAI tools like ChatGPT alongside traditional materials, emphasising the importance of consulting multiple sources and critically evaluating information.

Implications

The study findings offer several implications for the integration of GenAI technologies in individual formative e-assessment. Firstly, the varied success and challenges students face when interacting with ChatGPT underscore the need for AI literacy programmes. These programmes would aim to equip students with the skills necessary to effectively use AI tools, including understanding their capabilities and limitations, how to interpret AI-generated responses, assess their reliability and how to formulate queries to obtain useful responses.

Secondly, the strategic use of ChatGPT for assistance with formative e-assessments, including direct copying of questions into the tool, highlights the need for clear guidelines on the ethical use of AI in academic settings. Educational policies should clearly define what constitutes ethical AI use in coursework and formative e-assessments, providing examples of acceptable and unacceptable practices. This would help maintain academic integrity while allowing students to benefit from AI technologies.

Thirdly, the effectiveness of ChatGPT in providing explanations and verifying students' thought processes suggests that GenAI tools can play a complementary role in learning and assessment. However, the findings also suggest that traditional study methods and resources (e.g., textbooks, class notes) still remain crucial in supporting students' learning. Educators should consider how GenAI tools can be integrated with traditional teaching methods to enhance learning outcomes. This could involve using AI to support flipped classroom models, personalised learning, and as a supplementary resource for explaining complex concepts.

Lastly, the strategic use of ChatGPT for verification and exploration emphasises the importance of adaptive learning strategies and critical thinking skills. Educators should encourage students to assess the reliability of AI information, reflect on answers, and determine how best to incorporate AI tools into their learning. Additionally, formative e-assessments could be designed to not only test knowledge but also

engage higher-order thinking skills, including analysing and critiquing AI-generated content. This could include open-book formative e-assessments that assess higher-order thinking skills or the inclusion of questions that require students to critique or analyse AI-generated content.

Limitations and future work

This study is not without limitations. Primarily, the formative e-assessments lacked open-ended questions, focusing instead on interaction patterns with GenAI tools through multiple-choice and true/false questions. This approach standardised the assessment framework but did not fully assess critical thinking or synthesis skills. Future research should incorporate open-ended questions to explore how students use GenAI to develop more complex responses.

Apart from that, the study's scope was confined to a specific group of undergraduate students within a single academic discipline, which may limit the generalisability of the findings. Expanding future studies to include diverse academic fields, educational levels and cultural backgrounds could enhance the applicability of the results.

Moreover, the research also did not deeply explore the cognitive processes students use when interacting with GenAI tools. Qualitative methods such as interviews and focus groups could provide insights into the students' thinking, learning strategies, and decision-making.

In addition, focusing primarily on ChatGPT, the study does not explore the use or impact of other AI tools that might be employed for similar purposes in educational settings. Future studies could include a wider array of AI tools and technologies to compare and contrast their effectiveness, challenges and impacts on education. This would help educators and policymakers make informed decisions about integrating AI into learning environments.

Lastly, the findings suggests the need for ongoing research to enhance AI's accuracy and contextual sensitivity in educational contexts and to examine the effectiveness of AI literacy programmes, which could inform curriculum development and teaching strategies.

Conclusions

In conclusion, ChatGPT represents a significant frontier in the intersection of GenAI and education, offering opportunities to enhance learning experiences and assessment strategies. This study contributes to the ongoing dialogue on the role of GenAI in education, especially formative e-assessment. The frequency of students' GenAI use varied with question type, highlighting a preference for leveraging ChatGPT for multiple-choice and true/false questions. Moreover, the strategic use of ChatGPT for clarification, verification and exploration of concepts suggests that AI tools can complement traditional learning resources. In addition, although a significant portion of students adopted ChatGPT's responses, there was also a notable incidence of rejection due to various reasons, including misalignment with expectations, perceived inaccuracies and the tool's limitations in generating context-specific answers. This highlights the importance of integrating GenAI literacy into educational curricula, enabling students to effectively and ethically engage with GenAI tools. Moreover, the establishment of clear ethical guidelines for GenAI use in formative e-assessments is imperative to maintain academic integrity while harnessing the benefits of AI for educational enhancement.

Acknowledgements

This research was supported by a Future of Learning: AI Grant from Boston University Shipley Center. We gratefully acknowledge the support from our Shipley Center colleagues.

References

Alasadi, E. A., & Baiz, C. R. (2023). Generative AI in education and research: Opportunities, concerns, and solutions. *Journal of Chemical Education*, *100*(8), 2965–2971. https://doi.org/10.1021/acs.jchemed.3c00323

Al-Hattami, A. A. (2020). E-assessment of students' performance during the e-teaching and learning. *International Journal of Advanced Science and Technology*, *29*(8), 1537–1547. http://sersc.org/journals/index.php/IJAST/article/view/12566

Bahati, B., Fors, U., Hansen, P., Nouri, J., & Mukama, E. (2019). Measuring learner satisfaction with formative e-assessment strategies. *International Journal of Emerging Technologies in Learning*, *14*(07), 61–79. https://doi.org/10.3991/ijet.v14i07.9120

BaiDoo-Anu, D., & Owusu Ansah, L. (2023). Education in the era of generative artificial intelligence (AI): Understanding the potential benefits of ChatGPT in promoting teaching and learning. *Journal of AI*, *7*(1), 52–62. https://doi.org/10.61969/jai.1337500

Bozkurt, A., Xiao, J., Lambert, S., Pazurek, A., Crompton, H., Koseoglu, S., Farrow, R., Bond, M., Nerantzi, C., Honeychurch, S., Bali, M., Dron, J., Mir, K., Stewart, B., Costello, E., Mason, J., Stracke, C. M., Romero-Hall, E., Koutropoulos, A., … Jandrić, P. (2023). Speculative futures on ChatGPT and generative artificial intelligence (AI): A collective reflection from the educational landscape. *Asian Journal of Distance Education*, *18*(1), 53–130. https://doi.org/10.5281/zenodo.7636568

Chan, C. K. Y., & Hu, W. (2023). Students' voices on generative AI: Perceptions, benefits, and challenges in higher education. *International Journal of Educational Technology in Higher Education*, *20*, Article 43. https://doi.org/10.1186/s41239-023-00411-8

Chien, Y.-C., Wu, T.-T., Lai, C.-H., & Huang, Y.-M. (2022). Investigation of the influence of artificial intelligence markup language-based LINE chatBot in contextual English learning. *Frontiers in Psychology*, *13*, Article 785752. https://doi.org/10.3389/fpsyg.2022.785752

Cohen, D., & Sasson, I. (2016). Online quizzes in a virtual learning environment as a tool for formative assessment. *JOTSE*, *6*(3), 188–208. https://doi.org/10.3926/jotse.217

Cooper, G. (2023). Examining science education in ChatGPT: An exploratory study of generative artificial intelligence. *Journal of Science Education and Technology*, *32*(3), 444–452. https://doi.org/10.1007/s10956-023-10039-y

Daly, C., Pachler, N., Mor, Y., & Mellar, H. (2010). Exploring formative e-assessment: Using case stories and design patterns. *Assessment & Evaluation in Higher Education*, *35*(5), 619–636. https://doi.org/10.1080/02602931003650052

Darwin, Rusdin, D., Mukminatien, N., Suryati, N., Laksmi, E. D., & Marzuki. (2024). Critical thinking in the AI era: An exploration of EFL students' perceptions, benefits, and limitations. *Cogent Education*, *11*(1), Article 2290342. https://doi.org/10.1080/2331186X.2023.2290342

García-Jiménez, E. (2015). La evaluación del aprendizaje: de la retroalimentación a la autorregulación. El papel de las tecnologías [Evaluation of learning: From feedback to self-regulation – The role of technologies]. *RELIEVE – Revista Electrónica De Investigación Y Evaluación Educativa*, *21*(2). https://doi.org/10.7203/relieve.21.2.7546

George, A. S. (2023). The potential of generative AI to reform graduate education. *Partners Universal International Research Journal*, *2*(4), 36–50. https://doi.org/10.5281/ZENODO.10421475

Haleem, A., Javaid, M., & Singh, R. P. (2022). An era of ChatGPT as a significant futuristic support tool: A study on features, abilities, and challenges. *BenchCouncil Transactions on Benchmarks, Standards and Evaluations*, *2*(4), Article 100089. https://doi.org/10.1016/j.tbench.2023.100089

Hettiarachchi, E., Mor, E., Huertas, M. A., & Guerrero-Roldán, A.-E. (2015). Introducing a formative e-assessment system to improve online learning experience and performance. *Journal of Universal Computer Science*, *21*(8), 1001–1021. https://doi.org/10.3217/JUCS-021-08-1001

Hmoud, M., Swaity, H., Hamad, N., Karram, O., & Daher, W. (2024). Higher education students' task motivation in the generative artificial intelligence context: The case of ChatGPT. *Information*, *15*(1), Article 1. https://doi.org/10.3390/info15010033

Hodges, C., & Ocak, C. (2023, August 30). Integrating generative AI into higher education: Considerations. *EDUCAUSE Review*. https://er.educause.edu/articles/2023/8/integrating-generative-ai-into-higher-education-considerations

Hou, I., Mettille, S., Man, O., Li, Z., Zastudil, C., & MacNeil, S. (2024). The effects of generative AI on computing students' help-seeking preferences. In N. Herbert & C. Seton (Eds.), *Proceedings of the 26th Australasian Computing Education Conference* (pp. 39–48). Association for Computing Machinery. https://doi.org/10.1145/3636243.3636248

Shannon, S. E. (2005). Three approaches to qualitative content analysis. *Qualitative Health Research*, *15*(9), 1277–1288. https://doi.org/10.1177/1049732305276687

Jiao, H. (2015). Enhancing students' engagement in learning through a formative e-assessment tool that motivates students to take action on feedback. *Australasian Journal of Engineering Education*, *20*(1), 9–18. https://doi.org/10.7158/D13-002.2015.20.1

Joshi, A., Virk, A., Saiyad, S., Mahajan, R., & Singh, T. (2020). Online assessment: Concept and applications. *Journal of Research in Medical Education & Ethics*, *10*(2), 49–59. https://doi.org/10.5958/2231-6728.2020.00015.3

Jovanovic, M., & Campbell, M. (2022). Generative artificial intelligence: Trends and prospects. *Computer*, *55*(10), 107–112. https://doi.org/10.1109/MC.2022.3192720

Kumar, J. A. (2021). Educational chatbots for project-based learning: Investigating learning outcomes for a team-based design course. *International Journal of Educational Technology in Higher Education*, *18*, Article 65. https://doi.org/10.1186/s41239-021-00302-w

Lee, Y.-F., Hwang, G.-J., & Chen, P.-Y. (2022). Impacts of an AI-based *cha*bot on college students' after-class review, academic performance, self-efficacy, learning attitude, and motivation. *Educational Technology Research and Development*, *70*(5), 1843–1865. https://doi.org/10.1007/s11423-022-10142-8

Lin, C.-J., & Mubarok, H. (2021). Learning analytics for investigating the mind map-guided AI chatbot approach in an EFL flipped speaking classroom. *Educational Technology & Society*, *24*(4), 16–35. https://drive.google.com/file/d/1y1mZZSn937YbB0_NwU2wakXZ1dTUeVf7/view

Lin, M. P.-C., & Chang, D. (2020). Enhancing post-secondary writers' writing skills with a chatbot: A mixed-method classroom study. *Educational Technology & Society*, *23*(1), 78–92. https://drive.google.com/file/d/13C8BgZa9YR9_oyoN0BTr2UgIFauIxJ-0/view

Liu, C.-C., Liao, M.-G., Chang, C.-H., & Lin, H.-M. (2022). An analysis of children' interaction with an AI chatbot and its impact on their interest in reading. *Computers & Education*, *189*, Article 104576. https://doi.org/10.1016/j.compedu.2022.104576

Lock, S. (2022, December 5). What is AI chatbot phenomenon ChatGPT and could it replace humans? *The Guardian*. https://www.theguardian.com/technology/2022/dec/05/what-is-ai-chatbot-phenomenon-chatgpt-and-could-it-replace-humans

Luckin, R., Holmes, W., Griffiths, M., & Corcier, L. B. (2016). *Intelligence unleashed. An argument for AI in education*. Pearson.

Marimuthu, F., & Ramraj, U. (2019). An authentic e-assessment task. In K.-H. Huarng (Chair), *Proceedings of the 2019 International Conference on E-Business and E-Commerce Engineering* (pp. 41–46). Association for Computing Machinery. https://doi.org/10.1145/3385061.3385062

Marr, B. (2024, February 13). AI showdown: ChatGPT Vs. Google's Gemini – which reigns supreme? *Forbes*. https://www.forbes.com/sites/bernardmarr/2024/02/13/ai-showdown-chatgpt-vs-googles-gemini--which-reigns-supreme/

McCallum, S., & Milner, M. M. (2021). 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

McDonald, N., Johri, A., Ali, A., & Hingle, A. (2024). *Generative artificial intelligence in higher education: Evidence from an analysis of institutional policies and guidelines*. arXiv. https://doi.org/10.48550/ARXIV.2402.01659

Mohamed, N. (2020). Relationship between formative assessments given on Moodle and self-regulatory learning skills among blended learning students at the Maldives National University. In M. Barajas Frutos & S. Al-Majeed (Chairs), *Proceedings of the 2020 3rd International Conference on Education Technology Management* (pp. 48–54). Association for Computing Machinery. https://doi.org/10.1145/3446590.3446598

O'Connor, C., & Joffe, H. (2020). Intercoder reliability in qualitative research: Debates and practical guidelines. *International Journal of Qualitative Methods*, *19*. https://doi.org/10.1177/1609406919899220

Pack, A., & Maloney, J. (2023). Using generative artificial intelligence for language education research: Insights from using OpenAI 's ChatGPT. *TESOL Quarterly*, *57*(4), 1571–1582. https://doi.org/10.1002/tesq.3253

Raffo, D. (2024, June 10). Gemini vs. ChatGPT: What's the difference? *TechTarget*. https://www.techtarget.com/searchenterpriseai/tip/Gemini-vs-ChatGPT-Whats-the-difference

Stödberg, U. (2012). A research review of e-assessment. *Assessment & Evaluation in Higher Education*, *37*(5), 591–604. https://doi.org/10.1080/02602938.2011.557496

Voss, E., Cushing, S. T., Ockey, G. J., & Yan, X. (2023). The use of assistive technologies including generative AI by test takers in language assessment: A debate of theory and practice. *Language Assessment Quarterly*, *20*(4–5), 520–532. https://doi.org/10.1080/15434303.2023.2288256

Walker, D. J., Topping, K., & Rodrigues, S. (2008). Student reflections on formative e-assessment: Expectations and perceptions. *Learning, Media and Technology*, *33*(3), 221–234. https://doi.org/10.1080/17439880802324178

Westfall, C. (2023, November 16). New research shows ChatGPT reigns supreme In AI tool sector. *Forbes*. https://www.forbes.com/sites/chriswestfall/2023/11/16/new-research-shows-chatgpt-reigns-supreme-in-ai-tool-sector/

Wu, R., & Yu, Z. (2024). Do AI chatbots improve students learning outcomes? Evidence from a meta-analysis. *British Journal of Educational Technology*, *55*(1), 10–33. https://doi.org/10.1111/bjet.13334

Yilmaz, R., & Yilmaz, F. G. K. (2023). The effect of generative artificial intelligence (AI)-based tool use on students' computational thinking skills, programming self-efficacy and motivation. *Computers and Education: Artificial Intelligence*, *4*, Article 100147. https://doi.org/10.1016/j.caeai.2023.100147

Yin, J., Goh, T.-T., Yang, B., & Xiaobin, Y. (2021). Conversation technology with micro-learning: The impact of chatbot-based learning on students' learning motivation and performance. *Journal of Educational Computing Research*, *59*(1), 154–177. https://doi.org/10.1177/0735633120952067

Corresponding author: Dongpeng Huang, dphuang@bu.edu

Please cite as: Huang, D., Huang, Y., & Cummings, J. J. (2024). Exploring the integration and utilisation of generative AI in formative e-assessments: A case study in higher education. *Australasian Journal of Educational Technology*, *40* (4), 1–19. https://doi.org/10.14742/ajet.9467