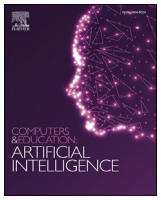
Computers and Education: Artificial Intelligence 7 (2024) 100269

Contents lists available at ScienceDirect

Computers and Education: Artificial Intelligence

journal homepage: www.sciencedirect.com/journal/computers-and-education-artificial-intelligence



Perceived impact of generative AI on assessments: Comparing educator and student perspectives in Australia, Cyprus, and the United States ☆

René F. Kizilcec [a,*], Elaine Huber [b], Elena C. Papanastasiou [c], Andrew Cram [b], Christos A. Makridis [d,e], Adele Smolansky [a], Sandris Zeivots [b], Corina Raduescu [b]

[a] Cornell University, Ithaca, NY, USA
[b] The University of Sydney Business School, Sydney, Australia
[c] University of Nicosia, School of Education, Nicosia, Cyprus
[d] Arizona State University, W. P. Carey School of Business, Tempe, AZ, USA
[e] University of Nicosia, Institute for the Future, Nicosia, Cyprus

ARTICLE INFO

Keywords:
Assessment
Generative AI
ChatGPT
Educators
Students
Survey

ABSTRACT

The growing use of generative AI tools built on large language models (LLMs) calls the sustainability of traditional assessment practices into question. Tools like OpenAI's ChatGPT can generate eloquent essays on any topic and in any language, write code in various programming languages, and ace most standardized tests, all within seconds. We conducted an international survey of educators and students in higher education to understand and compare their perspectives on the impact of generative AI across various assessment scenarios, building on an established framework for examining the quality of online assessments along six dimensions. Across three universities, 680 students and 87 educators, who moderately use generative AI, consider essay and coding assessments to be most impacted. Educators strongly prefer assessments that are adapted to assume the use of AI and encourage critical thinking, while students' reactions are mixed, in part due to concerns about a loss of creativity. The findings show the importance of engaging educators and students in assessment reform efforts to focus on the process of learning over its outputs, alongside higher-order thinking and authentic applications.

1. Introduction

In a remarkable convergence of research and real-world impact, the sudden emergence of ChatGPT has sent shockwaves through the global landscape of education. As students, educators, and university administrators grapple with the practical implications of generative AI, it becomes abundantly clear that we stand at the precipice of a new era. Since the release of GPT-3, a groundbreaking large language model (LLM) released by OpenAI, and its offspring, the user-friendly ChatGPT conversational interface, researchers are both excited and filled with trepidation over its boundless possibilities and transformative potential (Cotton et al., 2023; Farazouli et al., 2023; Nikolic et al., 2023). Generative AI, as defined by Weng (2023), refers to a technology that utilizes deep learning models to generate content that closely resembles human expression in response to complex and diverse prompts. These tools have the ability to produce conversational-style text that closely resembles human writing, as well as other visual and auditory media. They can be used to create systems that operate in ways that resemble human cognition and behavior (Siemens et al., 2022; Markel et al., 2023; Park et al., 2023). For example, ChatGPT and its derivatives are increasingly utilized for language translation, human-like conversation with chatbots, writing articles, stories, computer code, and other forms of written content (Cotton et al., 2023).

Generative AI tools promise many benefits in education, such as increasing student engagement in learning tasks, providing timely feedback, aiding research and collaboration, and improving accessibility (Kasneci et al., 2023). For example, AI technology can provide immediate feedback via automated grading (Mate & Weidenhofer, 2022) and

☆ This work was supported by a Global Strategic Collaboration Award between Cornell University and The University of Sydney, and by the National Science Foundation under Grant No. 2237593.

* Corresponding author.
E-mail addresses: kizilcec@cornell.edu (R.F. Kizilcec), elaine.huber@sydney.edu.au (E. Huber), papanastasiou.e@unic.ac.cy (E.C. Papanastasiou), andrew.cram@sydney.edu.au (A. Cram), cmakridi@stanford.edu (C.A. Makridis), as2953@cornell.edu (A. Smolansky), sandris.zeivots@sydney.edu.au (S. Zeivots), corina.raduescu@sydney.edu.au (C. Raduescu).

https://doi.org/10.1016/j.caeai.2024.100269
Received 30 November 2023; Received in revised form 15 July 2024; Accepted 17 July 2024
Available online 26 July 2024

facilitate the provision of meaningful feedback in large cohorts (Bernius et al., 2022). At the same time, AI raises serious concerns about the validity of widely used assessment practices, especially concerns about academic integrity and bypassing important learning processes (Swiecki et al., 2022). Because standard assessment practices focus on evaluating the final products like essays to measure learning, researchers have highlighted the potential for plagiarism as a key challenge with using ChatGPT for assessment in higher education (Cotton et al., 2023). Students can potentially use generative AI tools like ChatGPT to cheat on online assessments by submitting essays that are not their own work. The problem might be more prevalent in online assessments where students tend to feel more distant from their instructors (Papanastasiou & Solomonidou, 2023).

Educators can face challenges distinguishing between students' own work and responses generated by AI tools, making it difficult to assess students' level of understanding and their ability to apply the material (Mao et al., 2024). Unless educators and academic institutions adapt to this new reality, generative AI can undermine academic integrity in online assessments and the purpose of higher education to educate students, which may reduce the signaling effects and inherent value in formal educational attainment (Cotton et al., 2023). To address this major problem, scholars have called for applying AI in classrooms in such a way that promotes self-regulated and more productive learning, rather than treating it as a replacement for human effort in the learning process (Hopfenbeck et al., 2023; Mao et al., 2024; Swiecki et al., 2022).

AI has been framed as a transformative resource that educators and students can leverage in teaching and learning. Weng (2023) suggests ways to employ generative AI tools such as raising awareness of these tools, using them in class, in assessments, and engaging in discussions with students about their promises and challenges. They argue that this is more productive than either banning them or giving them a central role in the curriculum. Integrating generative AI with assessments can also transform assessment practices and experiences, for example, by immersing students in simulated learning environments where they can safely and repeatedly practice skills (Markel et al., 2023). This paradigm shift may require the development of new assessment approaches and policies that achieve a balance between the advantages of AI and the imperative to maintain academic integrity (Chan & Chen, 2023). Bearman et al. (2023) argued that educational assessment practice has not kept up with the digital transformation. Students and educators require better guidance on how to engage in meaningful interactions with AI systems for the purpose of assessment (Viberg et al., 2024). These interactions would directly assess students' learning process, critical thinking, and evaluative skills, not just their knowledge and comprehension. To this end, we expect to see revised guidelines and recommendations for educational assessment policies, incorporating input from stakeholders involved in assessment design to address the two major questions around AI integration in education: 'what' to assess (Sabzalieva & Valentini, 2023) and 'how' to assess it (Chan & Chen, 2023).

There are many ongoing conversations around what types of assessments are needed given the capabilities of generative AI tools. Bearman and Luckin (2020) emphasize that machines lack the ability to define quality or establish standards, making it crucial to develop assessment designs that prioritize the distinctly human capacity for defining quality standards. This raises questions on assessment standards and a move towards more authentic, adaptive, and continuous assessment (Gašević et al., 2023). Adapting current assessment practices in response to the ubiquitous availability of generative AI tools is timely but also effortful. As AI continues to play a pivotal role in society, assessments need to be adapted to ensure that they assess students authentically and ethically. Assessment approaches that foster human expertise and judgment are primed to gain greater significance through digital technologies (Dann, 2014; Nieminen et al., 2023; Bearman & Luckin, 2020).

This moment presents a rare opportunity for real innovation in current assessment practices, because most commonly used assessments were not conceived with access to powerful generative AI tools in mind. To meet the moment, we need to understand educators' and students' perspectives on the issue to achieve sustainable advances in assessment practices. We are especially interested in how much the perspectives of these two stakeholders—educators and students—are in alignment to provide a common ground. This may vary across contexts shaped by the local pace of technological adoption, institutional characteristics, cultural differences, and linguistic variation in technological efficacy (i.e., generative AI tools may work better in English than in other languages such as Greek). Within this context, we pose the following three research questions: (1) Which types of assessments do educators and students consider to be most impacted by generative AI? (2) How do educators and students think that students will be using generative AI in completing assessments? (3) And what are their preferences and attitudes toward adapting assessments to incorporate generative AI? Answering these questions by building on an established framework for examining assessment quality for university online assessments is essential since such knowledge is needed to guide efforts to reform future assessment practices.

2. Background

2.1. Assessment in the era of generative AI

Assessments are essential for measuring what students have learned (Mislevy et al., 2003). Contemporary assessment approaches (e.g., assessment as learning; Torrance, 2007) require students to take an active role in their own learning, encouraging them to develop and apply capacities for self-regulation, critical thinking, and evaluative judgment (Gašević et al., 2022). This brings assessments into closer alignment with high-level learning outcomes, for instance, as articulated in Bloom's Taxonomy (Bloom et al., 1956). In his original taxonomy, Bloom placed the cognitive task of evaluation at the top of the pyramid (followed by synthesis, analysis, application, comprehension, and then knowledge): it is a complex and rich task that requires critical thinking and the integration of knowledge domains, but it is resource-intensive to assess and give effective feedback on using traditional methods (Krathwohl, 2002).

The advent of online or digital assessment in higher education predates the emergence of generative AI tools, yet it has continually evolved to incorporate technological advancements that promise to enhance the learning and assessment experience. The transition from traditional to digital assessment methods has been driven by the need for scalability, flexibility, and the potential to provide immediate feedback, among other benefits. In line with Vygotsky's emphasis on the learning journey over the destination (Vygotsky & Cole, 1978), there has been a shift towards assessing more of the process by which students arrive at a final product or solution (Gašević et al., 2022). However, this transition has also raised concerns regarding academic integrity, the authenticity of student work, and the adequacy of these methods in evaluating higher-order thinking skills (Sillat et al., 2021; Viberg et al., 2024; Gikandi et al., 2011).

Recent developments in generative AI, particularly tools like ChatGPT, have introduced new dimensions to the ongoing discourse on digital assessment (Swiecki et al., 2022). These tools possess the capability to generate human-like text, solve complex problems, produce code, and even simulate entire conversations, raising both opportunities and challenges for assessment practices in higher education. While generative AI tools can potentially support personalized learning and assist in creating more authentic, dynamic assessment tasks, they also pose significant risks related to academic dishonesty, educational equity, and the dilution of critical thinking and creativity in student work (Perkins, 2023; Gipps & Stobart, 2009; Kasneci et al., 2023; Yu, 2023). The literature on the integration of generative AI tools in educational assessment is emerging, with studies beginning to explore the implications of these technologies for teaching, learning, and assessment (Swiecki et al., 2022). While there are an increasing number of studies demonstrating the possibilities of generative AI, few thus far provide any evidence of real-world

Table 1
Assessment design dimensions with descriptions as provided in the survey instrument.

Dimension	Defined as the extent to which the assessment ...
Academic integrity	ensures security against cheating, impersonation, and other forms of inappropriate assistance
Student experience	enhances convenience and comfort for students, motivation, and concentration, minimizes stress and anxiety, and technical complication
Authenticity	has similar tasks to those performed in workplace or professional settings
Information integrity	reduces the likelihood of privacy breach (i.e., unauthorized access to student personal data, content students generated in their assessments)
Quality feedback	enables the provision of quality feedback (e.g., timely, multiple formats such as media, text, encourages the use of feedback towards later assessment)
Equity of access	enables flexible conditions to complete the assessment (e.g., ease of access for students with a disability/impairment, limited access to technology, geographically dispersed)

impacts for students and educators. Any impacts are going to be strongly shaped by what educators and students think about generative AI for assessments. Prior work shows that student and educator perspectives are critical enablers or barriers for innovation in educational assessments (Andrade & Du, 2019; Bevitt, 2015; Bennett et al., 2017).

Our study seeks to contribute to this burgeoning field by examining the perspectives of key stakeholders—educators and students—on the use of generative AI tools in assessment practices within higher education. By focusing on stakeholders' views, this research aims to shed light on the perceived benefits and drawbacks of generative AI in assessments, the types of assessments most impacted by these technologies, and the preferences and attitudes towards adapting assessment practices to incorporate generative AI. In doing so, we aim to better connect our findings to the existing body of research and to contribute valuable insights into the ongoing conversation about the future of assessment in an AI-augmented educational landscape.

2.2. Conceptual framework for online assessments

In a national study of Australian educators in the business disciplines, researchers investigated the factors that influence the design and evaluation of online assessments, which are defined as any type of graded activity with an online component designed to measure students' mastery of knowledge and/or skills (Huber et al., 2024). The authors identified six design dimensions and four contextual factors at play. The design dimensions include ensuring academic integrity, providing valuable feedback, creating a positive learning experience for students, delivering authentic assessment tasks, maintaining the integrity of student information, and ensuring equal opportunities for all students to complete the assessment successfully (Table 1). The broader contextual factors that influence assessment design decisions and practices include scale of delivery, resource constraints, institutional policies, and accreditation requirements. Their study also identified constraints and trade-offs that need to be negotiated in designing, evaluating, and implementing online assessments; the most important one as identified by educators, was academic integrity (see definition in Table 1). Online assessment presents practical challenges for student authentication and academic integrity: identity verification is especially difficult in online essays, and remote invigilation is challenging for online examinations (Mate & Weidenhofer, 2022; Cram et al., 2022).

Several of these challenges with designing and implementing new online assessments were highlighted in a recent study of business educators (Cram et al., 2022). They surveyed 97 university faculty members in Australia about their views on assessments that are not conducted with pen and paper, but digitally through technology. Their focus was not on AI but they considered challenges that arise in online assessments that also apply to working with AI tools. These challenges include additional academic teacher time and effort, the logistics and timing of new kinds of assessments, technology access, consistency over time, functionality and usability, alignment with student preferences and expectations, effectively preparing students for new assessment formats, and institutional and departmental policies that might inhibit new assessment designs and implementations. Considering these challenges, it is important to understand educators' attitudes towards generative AI in the context of assessments. Yet, educators are not the only ones who will be quick to respond to assessment reform efforts. Students have been looking for guidance on how to approach assessments that were not designed with the capabilities and availability of modern AI tools in mind. Moreover, both educators and students look to school leadership to provide official guidance on academic integrity and effective uses of AI in teaching, learning, and assessment. In Australia, for example, the Tertiary Education Quality and Standards Agency (TEQSA) has compiled a set of resources to assist higher education providers in addressing the challenges and leveraging the opportunities brought by the advances in generative AI tools (Tertiary Education Quality and Standards Agency, 2024). This has prompted universities to embed guiding statements about the use of AI tools in their policies.

In this research study, we build on the (Huber et al., 2024) framework, which has thus far been tested in the Australian higher education context with educators, by testing it as a heuristic to evaluate the impact of generative AI on assessments in new contexts. We collected data in two other continents, North America and Europe, to examine what similarities and differences might appear across the three international institutions, and thus acknowledge the value of understanding these dynamics in varied educational and cultural settings. We are also interested in students' perspectives on the design of online assessments and particularly the impact of generative AI tools. This study centers around both, the educator and student perspective to gain a holistic and comparative understanding of how these two groups make sense of how recent technological changes affect assessments.

3. Methods

3.1. Participants and contexts

We collected survey data from students and educators at three institutions of higher education located in three countries across three continents (Australia, Cyprus, and the United States). The sample characteristics are shown in Table 2. Participants were recruited through email to complete a 15-minute survey through Qualtrics (Australia, USA) or SurveyMonkey (Cyprus). For the US sample, incentives to respond included a $5 charity donation and course credit for students; no incentives were offered for Australian or Cypriot respondents. Respondents who provided informed consent at the start of the survey and answered the first page of questions were retained for analysis. Given that each survey sample was collected in a specific institution of higher education, which is not representative of each country, we provide further information on the institutional context. The convenience samples are also not representative of the institutions where they were collected.

In Australia, we collected responses at The University of Sydney, a large urban, public university classified as research-intensive, with around 70,000 students and 7,000 staff. The student survey was distributed by an announcement in the university Learning Management System and the faculty survey was advertised on internal social media channels and the various Faculty and School staff mailing lists.

In Cyprus, we collected responses at the University of Nicosia, a private urban university in the Republic of Cyprus. In the distribution of European countries, Cyprus ranks below the average size and gross domestic product. The institution enrolls around 14,000 students and has

Table 2
Survey sample characteristics for students and educators by country.

	Australian Sample		Cypriot Sample		US Sample	
	Students	Educators	Students	Educators	Students	Educators
Sample Size	353	29	276	48	51	10
Age Mean (SD)	24.1 (7.9)	–	30.7 (6.7)	45.5 (7.7)	21.1 (1.1)	48.2 (16.9)
Undergraduates	48%	–	14%	–	100%	–
Graduates (Master/PhD)	52%	–	86%	–	0%	–
Yrs. Teaching Mean (SD)	–	13 (9)	–	15 (8)	–	18 (19)
Discipline						
Agriculture, Life, Health	0%	0%	5%	6%	58%	25%
Arts and Science	14%	54%	7%	33%	11%	0%
Business	32%	38%	2%	11%	22%	0%
Education	0%	0%	78%	44%	0%	0%
Engineering	50%	8%	6%	6%	8%	50%
Law	3%	0%	1%	0%	0%	25%
Medicine	0%	1%	0%	0%	0%	0%

Table 3
Original and adapted assessment prompt for each assessment type.

Essay Assessment Prompt
Original: Write a 5-page essay on [a given topic in your discipline; e.g., Greek mythology, human rights, sustainable energy, sorting algorithms]. You have 7 days to complete the essay.
Adapted: You are given a 5-page essay produced by ChatGPT on [a given topic in your discipline; e.g., Greek mythology, human rights, sustainable energy, sorting algorithms]. You have 7 days to analyze the essay and edit it yourself to improve its quality, making clear references to the original text where applicable.
Coding Assessment Prompt
Original: 1. Write two different algorithms using Python code to sort a list of numbers. 2. Evaluate the correctness of each approach (1–2 paragraphs each). 3. Analyze the time complexities (i.e., how long they take depending on the length of the list) of each algorithm (1–2 paragraphs each). You have 7 days to submit your solution to the above questions.
Adapted: Evaluate and compare two different algorithmic approaches to sort a list of numbers by following these three steps: 1. Ask ChatGPT to generate an algorithm using Python code to sort a list of numbers. Provide the output. In 1–2 paragraphs, explain whether you think the code is correct. Include examples of using the algorithm. 2. Ask ChatGPT to generate a more efficient algorithm using Python code to sort a list of numbers. Provide the output. In 1–2 paragraphs, explain whether you think the code is correct. Include examples of using the algorithm. 3. Ask ChatGPT to analyze the time complexity of the algorithms (i.e., how long they take depending on the length of the list). Provide the output. In 1–2 paragraphs, explain whether you agree with the output. You have 7 days to submit responses to the above questions.

about 1,100 full-time and part-time faculty and staff. The survey announcement was distributed to faculty members via e-mail, and to the students through the university's Learning Management System.

In the US, we collected responses at Cornell University, a private land-grant university located in a rural region of the Northeastern part of the country. It enrolls around 15,000 undergraduate and 10,000 graduate students and has 2,000 full-time faculty. The student survey was distributed via course email lists, flyers, and social media postings. The faculty survey was sent to a large email list of faculty with an interest in technology and related fields.

3.2. Measures

The survey included the same set of questions for students and educators. Our goal from the start was to understand how much overlap there is in the views of educators and students. To achieve a direct comparison, we hold the questions constant between the two stakeholders. This way we can reach internally valid claims about the similarities and differences of their attitudes at the same point in time, unaffected by variation in how each question was asked.

3.2.1. Awareness and use of generative AI

We measured awareness with a single yes/no question: "Have you heard of ChatGPT before?" To ensure respondents understood the question consistently, we included a short definition above it: "Generative AI tools use artificial intelligence (AI) to generate new content such as text, imagery, audio and synthetic data. ChatGPT is an example of a generative AI tool that uses natural language processing to generate text. ChatGPT can be used to prompt graphics, create articles, write product descriptions, explain complex topics, and more."

We measured generative AI use using a Likert-style question: "How frequently do you use ChatGPT (or similar generative AI tools) in the following ways?" The following categories were provided: coursework; research; non-academic, professional purposes; and fun. Respondents rated their frequency of use for each category on the following scale: Never, A few times, Weekly, Daily. We aggregate the number of weekly and daily responses.

3.2.2. Perceived impact of generative AI on assessment types

The perceived impact of generative AI on different types of assessments was measured using a Likert-style question: "How much do you expect each type of assessment to be impacted by the introduction of generative AI tools like ChatGPT?" For eight different types of assessments (see Fig. 1), respondents rated the perceived impact on the following scale: Not at all, Slightly, Moderately, Very, Extremely, Unsure. We excluded "Unsure" responses in the analysis.

3.2.3. Expected student use of generative AI

We asked an open-ended question to understand how students and educators thought that students would use generative AI to complete two specific assessments. After viewing each of the original assessment prompts (see Table 3), and before seeing the adapted prompt, participants were asked: "How do you think students might use generative AI tools to complete this assessment?" We coded these open-ended responses and identified themes using the method explained in the analytic approach.

3.2.4. Preference for assessment adaptation

We used a scenario-based design to assess preferences with two different assessment scenarios (essay and coding). Table 3 shows the assessment prompts that were used. After seeing both the original and adapted assessment prompt for each context, participants were asked to choose between the original prompt, adapted prompt, or indicate no preference between the two: "Assuming ChatGPT is available, which

Table 4
Awareness and uses of ChatGPT among students and educators by country.

Sample	Awareness (%)	Use Daily/Weekly for (%)			
		Coursework	Research	Professional Work	Fun
Educators	100	7	20	28	22
Australia	100	7	17	31	28
Cyprus	100	9	20	30	15
USA	100	0	30	10	40
Students	74	19	18	20	16
Australia	97	29	28	33	25
Cyprus	42	6	5	5	6
USA	98	24	14	12	14

assessment do you prefer?" We also asked educators to predict what students would prefer, and vice versa: "Assuming ChatGPT is available, which scenario do you think students (educators) prefer?"

3.2.5. Understanding assessment adaptation preferences

After indicating their assessment preference for each assessment type (first essay and then coding), participants were asked: "Please explain why you chose the preference responses above." We coded these open-ended responses and identified themes using the method explained in the analytic approach.

Moreover, to unpack assessment adaptation preferences quantitatively, participants rated each assessment using the established dimensions in Table 1 (Huber et al., 2024) to understand how generative AI and assessment adaptation were perceived to affect assessment quality. First, participants read the original prompt and answered: "Consider the assessment prompt prior to the availability of generative AI tools, such as ChatGPT. To what extent does this assessment ensure each design assessment dimension?" Participants rated the dimensions on a 5-point scale: None, Low, Medium, High, or Not Sure. We excluded "Not Sure" in the analysis. Then, participants were instructed, "Now consider the same assessment prompt again but in today's context when generative AI tools, such as ChatGPT, are widely available (i.e., students have equal access to the tool). To what extent does this assessment ensure each design assessment dimension?" They once again rated the dimensions in Table 1 on the same scale. Finally, participants were shown the adapted version of the assessment prompt and instructed to "consider an adapted version of the assessment prompt, given today when generative AI tools, such as ChatGPT, are widely available. To what extent does this assessment ensure each design assessment dimension?" For a third time, participants rated the dimensions on the same scale.

3.2.6. Perceived importance of assessment quality dimensions

We measured the perceived importance of the assessment dimensions in Table 1 using a Likert-style question: "How important for the quality of an assessment do you consider each of these dimensions to be?" Respondents rated each dimension on a 5-point scale from 'Not at all important' to 'Extremely important'.

3.3. Analytic approach

We utilized a multi-stage, mixed-method analytic approach that combined quantitative and qualitative analyses. The quantitative data from this study were analyzed using common descriptive, correlational, and regression approaches in *R*. The focus of these analyses was to examine the differences between educators and students and, in some cases, to examine variations between respondents in different countries.

For the qualitative analysis of open-ended responses, we used a thematic coding approach for each question. We developed codebooks with the help of GPT-4, carefully validated by the research team, and then manually coded responses. Recent work has tested the efficacy of GPT-4 for qualitative coding and found positive results (Zambrano et al., 2023). We prompted ChatGPT with the Advanced Data Analysis Beta plugin to find themes in the responses that were uploaded. We also provided context about the research design and goals in the prompt. We requested that each theme be supplemented by examples from the responses. To check the reliability of themes across regions and repetitions, we had researchers in Australia, Cyprus, and the US conduct the thematic prompting three times each.[1]

We aggregated the results, which differed only slightly, to obtain a summary of themes. Due to data use restrictions, we could only use GPT with responses from Cyprus and the US. For the Australian responses, two independent members of the research team who were unaware of the GPT-aided codebook manually created a thematic codebook from scratch. The manual coding process resulted in a codebook that was well-aligned with the machine-generated themes: for the two open-ended questions that were analyzed in this manner, the human raters identified 5 out of 6 and 6 out of 6 themes from the GPT-aided codebook. We used all six themes for each question for the final manual coding of the responses.

4. Results

4.1. Awareness and uses of generative AI

To position and compare the three samples we collected in terms of their familiarity with generative AI, Table 4 shows how aware educators and students were of ChatGPT, and how they tended to use it. We found that educators and students in our sample were generally aware of ChatGPT (or similar generative AI tools), except for students in Cyprus, who reported lower levels of awareness (42%; $\chi^2_2 = 262, p < 0.001$). We also found that a remarkable number of educators in our sample were using ChatGPT or similar tools frequently for research, professional work, and fun. Many of the students in Australia and the US, but not Cyprus, were also regularly using ChatGPT for coursework, research, professional work, and fun.

Students' awareness and aggregate use of ChatGPT were not associated with their age, degree type, or the year in their program (awareness: $F_{10,247} = 0.68,\ p = 0.74$; use: $F_{10,244} = 0.51,\ p = 0.88$). Likewise, educators' awareness and aggregate use of ChatGPT were not associated with how many years they had been teaching (awareness: $F_{1,28} = 0,\ p > 0.9$; use: $F_{1,27} = 2.20,\ p = 0.15$).

4.2. Expected impact of generative AI on assessments

Students and educators across the three samples reported similar expectations about the impact that generative AI tools on different types of assessment (Fig. 1). Educators generally rated coding assignments and essay prompts as the two most impacted assessments. These were also rated as strongly impacted by students along with short-answer and multiple-choice questions. The types of assessment expected to be most impacted were the ones that require written inputs and have a clear (set

[1] Cyprus ranks below the average size and gross domestic product in the distribution of European countries. As a result, we urge caution when interpreting and comparing across our country samples.

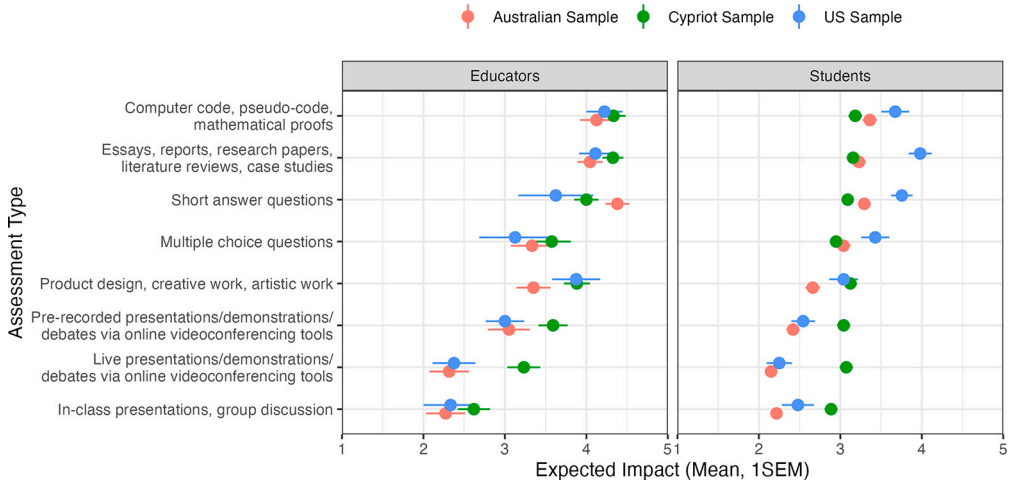


Fig. 1. Mean ratings of how much different types of assessment are perceived to be impacted by the availability of generative AI tools like ChatGPT according to educators (left panel) and students (right) in each country (color). The scale points are labeled *Not at all* (1), *Slightly* (2), *Moderately* (3), *Very* (4), and *Extremely* (5). Standard error bars are shown.

Table 5
Expected uses of generative AI: thematic analysis results with relative frequency of occurrence among educators and students.

Themes for Essay	Educators	Students	Themes for Coding	Educators	Students
Content Generation	57%	23%	Code Generation and Automation	62%	50%
Research and Idea Generation	17%	33%	Guidance, Validation, Error-Checking	21%	22%
Rewording and Revision	11%	19%			
Academic Integrity/Ethics	6%	6%	Academic Integrity/Ethics	3%	4%
Learning/Thinking Inhibitor	6%	2%	Learning/Thinking Inhibitor	–	5%
Learning Enhancement	4%	12%	Learning Enhancement	14%	13%
Efficiency and Time-saving	–	5%	Efficiency and Time-saving	–	6%

of) correct solution(s): for example, short answers to questions, mathematical proofs, computer code, and multiple-choice answers. In contrast, educators and students expected assignments that require real-time presentation (in-person or online) to be impacted the least.

Students in the Cypriot sample appeared to be unsure how assessments would be impacted, rating all of them as "moderately impacted" (3 on the 1–5 scale) in Fig. 1. This result may be related to their lower level of awareness and engagement with ChatGPT.

4.3. Expected student use of generative AI in assessments

Given that both educators and students in our sample believed that essay-style and coding-style assignments are most impacted by generative AI, we examined how they thought that students would use tools like ChatGPT to complete assessments. In particular, educators and students wrote about how they expected students might use generative AI to complete the original version of each prompt in Table 3. Thematic coding of all open-ended responses revealed a set of expected approaches that did not vary much based on the assessment (essay and coding) and respondent group (educators and students). Table 5 shows the identified themes and their relative frequency.

The identified themes covered a range of expectations, but also opportunities and concerns. While many educators and students expected students to use tools like ChatGPT to write an entire essay or write all of the code, this inclination was weaker among students who were more likely to think that their peers would use it to get help in the process (e.g., research and idea generation; rewording and revision; guidance, validation, error-checking). For example, a student in Cyprus wrote: *"Students in order to complete this assessment might use AI tools to help them check their knowledge, detect and solve errors and finally give them the right feedback."* Another student in the US wrote: *"Ask it to write something, and then modify it. Ask it to teach the user about the topic. Maybe ask it to make an essay better."*

Academic integrity concerns were brought up a few times by both educators and students, but only students mentioned expected efficiency gains from using generative AI. For example, a student in Australia wrote: *"Finding obscure resources quicker, reducing manual searching time. This is huge as many students work and study."* Both groups noted an opportunity for enhancing learning and only a small number raised concerns about interfering with learning and critical thinking. The exception was educators commenting on the essay assessment, who raised more concerns about learning and mostly expected students to use ChatGPT to generate the essay content.

4.4. Preference for assessment adaptation

After respondents reported how students might use generative AI to answer the original prompt, we showed them an adapted version of the same assessment that intentionally incorporated generative AI into the process (see Table 3). We asked for their preference between the original and the adapted prompt (including an option for expressing indifference) for each assessment (essay and coding). The preference rating results for the full sample are shown in Fig. 2. Educators strongly preferred the adapted prompt: 67% would select it, while only 16% would select the original prompt ($\chi^2 = 40.7, p < 0.001$), and 17% were undecided. Students also preferred the adapted prompt but not as strongly: 41% would select the adapted prompt, while 33% would select the original ($\chi^2 = 7.67, p = 0.006$), and 26% were undecided. We also found adaptation preferences to be remarkably similar between the two types of assessment ($z = -0.480, p = 0.631$), even though the assessments and their adaptation were distinct.

In addition, we asked educators and students to predict what they thought the preference of the other group would be (i.e., what educators thought students would prefer, and vice versa). This yielded an almost identical response to what is shown in Fig. 2, with one exception: for the essay assessment, 37% of educators thought that students would prefer

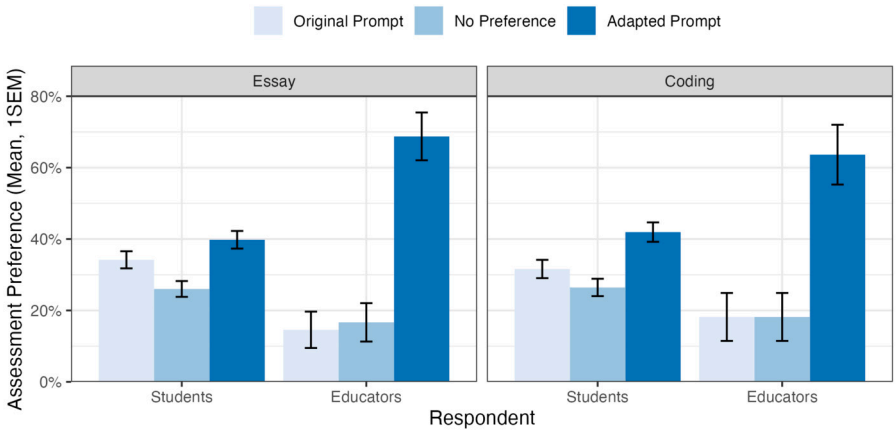


Fig. 2. Percentage of respondents who preferred the adapted assessment prompt, which incorporates generative AI into the task, themselves versus what they think the other stakeholder prefers, by assessment domain (panels). Standard error bars are shown.

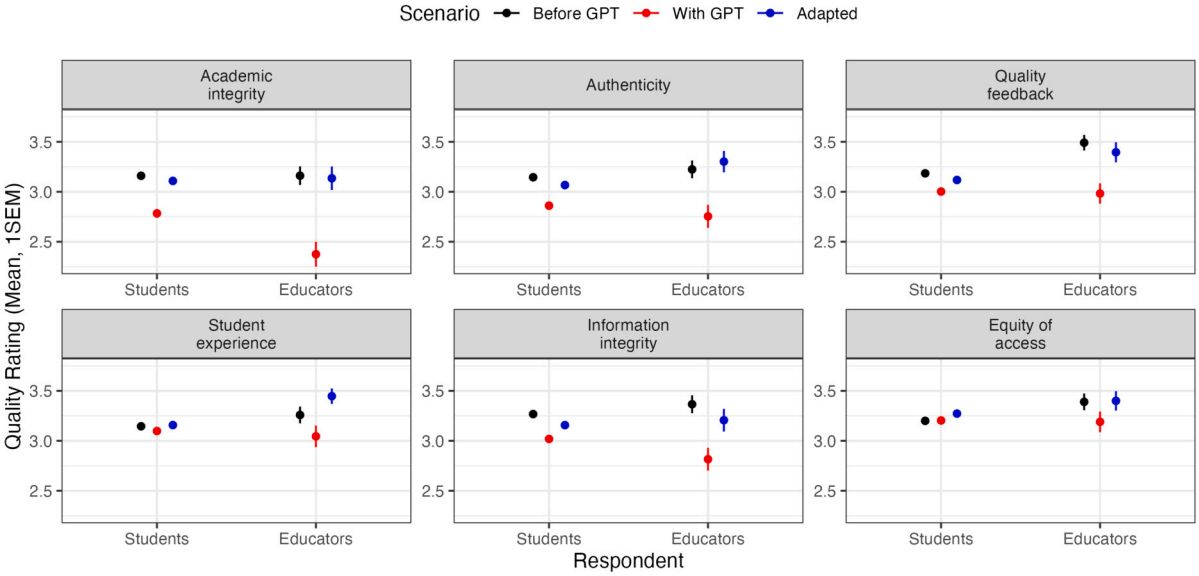


Fig. 3. Educator and student ratings of assessment prompts along six dimensions of assessment quality. Ratings show the perceived impact of generative AI (ChatGPT) on the quality of the original prompt (Before vs. With ChatGPT) and the impact of modifying the assessment prompt (With ChatGPT vs. Adapted). The scale points are labeled *None* (1), *Low* (2), *Medium* (3), and *High* (4). Standard error bars are shown (hidden under the dot for students).

the original prompt, while only 15% of educators preferred it themselves. Their prediction here is relatively accurate as 34% of student respondents preferred the original prompt. This pattern was observed for the essay but not the coding assessment.

4.5. Understanding assessment adaptation preferences

We used a mixed-methods approach to better understand the reasoning behind educators' and students' preferences for and against assessment adaptation. We examined ratings of the assessment prompts on each of the six dimensions of assessment quality (Table 1) to see how educators and students thought that adaptation affected the assessment. In addition, we thematically coded open-ended responses in which educators and students explained their preference ratings for each assessment.

Quantitative results: Fig. 3 shows educator and student ratings of the assessment prompts on each of the six dimensions of assessment quality. Ratings for the essay and coding assessments were combined in the visualization because they showed the same pattern of results. The results show a consistent and statistically significant ($p < 0.05$) pattern for most dimensions of assessment quality: the introduction of generative AI negatively affected ratings of the original prompt, but the adapted prompt recovered this loss in quality. This pattern was generally stronger among educator responses than student responses. Educators thought that the adapted prompt would provide an even better student experience than the original prompt before generative AI, while students rated the student experience to be similar.

Before we asked educators and students in our sample to judge the specific assignments by rating them on the six dimensions of assessment quality, we checked if they considered the dimensions to be important. We found that all dimensions of assessment quality in the framework were considered to be, on average, "very important" (corresponding to 4 on a 1–5 scale) by educators and students. Fig. 4 shows the average ratings, which were consistently higher among educators than students ($t = 4.97, p < 0.001$). Educators rated academic integrity as slightly more important than the other dimensions ($t = 4.27, p < 0.001$). Students rated Quality Feedback slightly higher than both Information Integrity ($t = 2.91, p = 0.004$) and Student Experience ($t = 2.10, p = 0.036$). There was a small amount of variation between countries: students' importance ratings were a little lower in the Cypriot sample ($t = -4.04, p < 0.001$), while educators' importance ratings were a little lower in the US sample ($t = -2.42, p = 0.018$).

Qualitative results: Our thematic analysis of open-ended explanations separated responses based on the stated preference to understand reasons for that specific preference (original or adapted). Among those

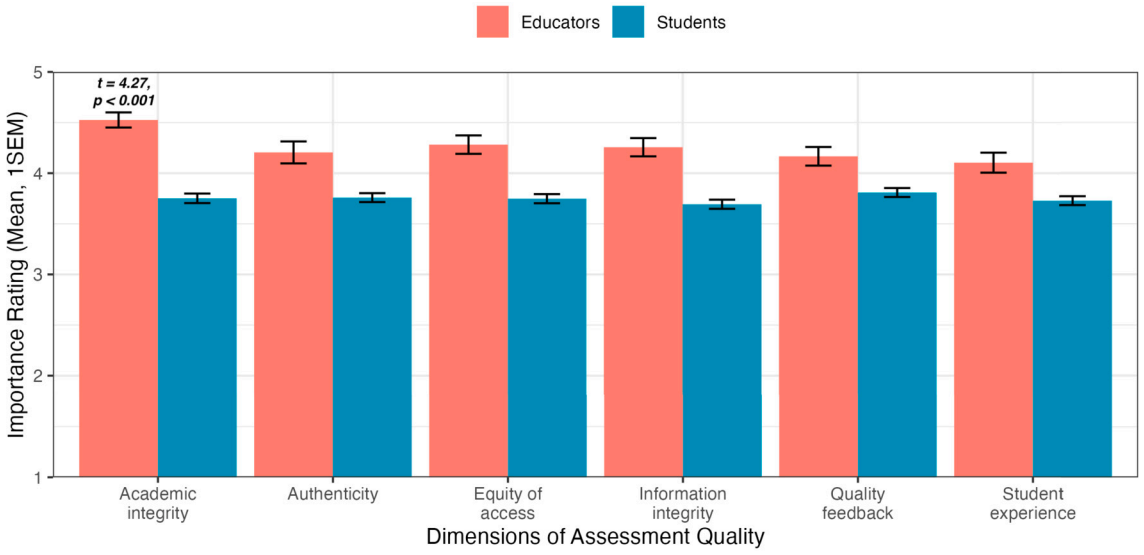


Fig. 4. Importance ratings for each dimension of assessment quality among educators and students. Standard error bars are shown.

who preferred the original assessment prompt, students were more inclined to explain what they disliked about the adapted prompt. For example, both students and educators expressed concerns about academic integrity as a reason for preferring the original prompt, with one student writing: *"Improving an essay written by AI is by no means academic research"* (Student, Cyprus). Both students and educators noted convenience and the amount of time it takes as reasons for their preference for the original prompt, but they prioritized different elements. Students were more concerned with ease of completion (noting that they could use ChatGPT to complete the original prompt), while educators were more focused on the quality: e.g., *"The focus should be on enhancing the information and making quality improvements rather than generating content entirely"* (Educator, USA). Furthermore, both groups agreed on the importance of critical thinking, creativity, and original contributions in essay assessments. One student in the US sample wrote: *"The original prompt will still let me create something from scratch if I choose to not use ChatGPT. Thus, it is more authentic and I will have a learning experience."* Highlighting a student's original contribution is important either way, as an educator in the Cypriot sample remarks: *"I believe that a presentation of the essay is critical as educators will understand their comprehension of the AI-created information and the student's contribution to the quality of the essay."*

For those who preferred the adapted prompt, common themes among educators and students were that the adapted prompt provided more opportunities for cognitive skill development, as well as ease and efficiency. However, there were a few differences between educators' and students' comments. Educators noted that the adapted prompt may make learning more fun and motivating: e.g., *"They get to engage with a fun tool"* (Educator, Australia). Students cited a preference for improvement over creation as a reason for their adapted prompt preference: e.g., *"I prefer the adapted prompt, as a student, as it involves identifying weaknesses of written work and utilising MY OWN skills to refine the work into a superior piece"* (Student, Australia). Students also cited being interested in advances in education as a reason: e.g., *"As AI tools become more advanced, it's interesting to think about how to adapt the education system to enhance learning"* (Student, USA).

5. Discussion

In this study, we used a theoretically grounded instrument to understand educator and student attitudes and preferences about the impact of generative AI on assessments. We examined variation in attitudes across contexts, backgrounds, and prior experiences. Instead of finding variation between educators and students and across countries in line with varying educational approaches and cultural norms, we observed a surprising level of agreement across stakeholders and geographically distant institutions. An important exception is that students were more hesitant to endorse the adapted prompt than educators for reasons including a perceived loss in creativity and critical thinking. This highlights an important need for carefully designing and framing new assessment prompts in ways that center the process of learning, higher-order thinking, and authentic tasks.

Our research also speaks to the awareness of different groups to the opportunities of new technologies. Economics research on cross-country productivity suggests that differences in adoption and diffusion play an important role (Foster & Rosenzweig, 2010). Our results highlight how perceptions of generative AI—whether they are accurately informed or not—can play a role in the adoption of new technologies and practices.

5.1. Implications for theory

The implications of integrating tools like ChatGPT into higher education assessment practices encourage a re-evaluation of existing theoretical frameworks, particularly with regard to fostering higher-order thinking skills. As our findings have indicated, evaluation skills, which were once seldom explicitly cultivated in students and were even more rarely assessed, are now essential when interacting with LLMs. This prompts us as educators to embed evaluative components directly into revised assessments to facilitate the development of these critical skills. Bloom's revised taxonomy (Anderson et al., 2001) has long been sprucing this idea with its emphasis on higher-order cognitive skills and can serve as a foundational reference point for these developments. Alongside this, a focus on assisting students in developing skills in evaluative judgment (Tai et al., 2018) will gain further relevance. The new landscape of AI-infused education calls for a nuanced theory that delineates when and how AI-based adaptation of pedagogical and assessment strategies is warranted.

There is also an open question about how generative AI will affect productivity, and whether it may accelerate income inequality and polarization in the labor market (Frank et al., 2019; Acemoglu & Restrepo, 2018). Preliminary evidence from OpenAI indicates that 80% of the U.S. workforce could have at least 10% of their work tasks affected by the introduction of large language models (LLMs), and 19% of workers may see at least 50% of their tasks impacted (Eloundou et al., 2023). However, they also find that access to an LLM could accelerate the time it takes to complete 15% of all worker tasks, and 47–56% of all tasks when incorporating software and tooling built on top of LLMs. Concretely, in a recent randomized experiment with 5,179 customer support

agents, access to LLMs was found to increase their average productivity by 14% (Brynjolfsson et al., 2023).

Our study contributes to the emerging body of literature on the integration of generative AI tools in higher education assessment by providing empirical insights into the perspectives of educators and students. It underscores the need for further theoretical development to understand how generative AI can be aligned with educational objectives without compromising academic integrity and the assessment of higher-order thinking skills. In particular, generative AI could augment existing assessment frameworks but also calls for caution in ensuring that these technologies are used in ways that enhance, rather than diminish, educational outcomes.

5.2. Implications for practice

Our results highlight a general tendency for faculty to perceive a bigger impact of how assessments will change following the emergence of ChatGPT and a stronger preference for adapted assessment prompts compared to students. To the extent that LLMs might encourage greater student participation, or reduce barriers to engaging in learning activities, our results suggest that LLMs could improve the rate of human capital accumulation and influence the entire curriculum for university programs. For instance, generative AI could function as a personalized learning assistant that works every hour, every day, tailoring content and communication styles to the individual learner. As a result, learning could become more effective per hour spent.

LLMs might also serve as a catalyst for reforming university curricula and assessments. By embedding LLMs in the whole educational process, educators and students can strategically be steered away from lower-order cognitive skills, such as knowledge recall and comprehension, to focus more on higher-order skills, such as evaluation and critical thinking. This transition is not merely pedagogical but also pragmatic since helping students develop higher-order thinking skills, such as evaluating machine-produced content, is necessary for them to be able to navigate the complexities of professional and personal life in the 21st century. By embedding LLMs into the curriculum, educators can create learning environments that encourage active engagement, collaboration, critical thinking, problem-solving, and evaluation, which are all essential skills for the future of work.

The shift created by embedding LLMs into assessments could however engender a measure of resistance or discomfort among some students who might be unaccustomed to assessments that demand these adapted types of assessment tasks. For example, tasks requiring the evaluation of machine-generated content could be met with skepticism, as has been evident from the student's qualitative comments. It is crucial to recognize that this discomfort is less a critique of the method than it is a reflection of an educational system that has traditionally prioritized other forms of assessments. So since these higher-order skills are increasingly indispensable in contemporary work settings, integrating LLMs into both curricula and assessments, and especially assessments that measure critical and evaluative thinking skills (not just comprehension), becomes not just innovative but necessary. By doing so, educational institutions can better align learning outcomes with the real-world challenges and complexities that await students post-graduation in the labor market.

Our study offers insights into the current attitudes and preferences of educators and students regarding the use of generative AI in assessments. There is a general openness to adapting assessment practices to incorporate these technologies, albeit with concerns about academic integrity and the preservation of critical thinking. Based on our findings, we recommend that educational institutions consider pilot projects or experiments to explore the effective integration of generative AI in assessments, with a focus on transparency, student engagement, and the development of guidelines to maintain academic standards. However, we acknowledge that these recommendations are based on the perceptions and attitudes captured in our survey and that actual implementation should be approached with careful consideration of the specific institutional context, resources, and educational goals.

5.3. Limitations

Despite the novelty and significance of the current study, due to the fact that it provides an important early contribution to the literature on the use of and attitudes about generative AI, we also recognize some of its limitations. The first limitation pertains to the nature of the sampling methodology. Although we attempted to distribute the current survey as widely as possible, our samples were not random, as we purposely administered the questionnaires within our own institutions. The participants who chose to participate in the study did so voluntarily. It is therefore likely that response bias has been introduced in the data and that the respondents are not a truly representative sample from their corresponding populations. The sample is likely to over-represent higher-performing students and faculty who have a greater interest in generative AI and its applications. As a result, their responses might be upward biased and might convey different perspectives related to generative AI than non-responding participants.

The second limitation of the study lies in the time point at which the data were collected. This sample reflects attitudes from a cross-section of respondents during the spring semester of 2023, which reflects the early days of generative AI, whose uses and visibility are continuing to grow each day. The impact and applications of AI are continuing to grow rapidly each day and our results serve as a snapshot of AI attitudes in relation to assessment at a point in time. This can serve as a baseline that facilitates useful future comparisons. In light of this context, it would be interesting to replicate this study in several months, or at least another year, to gauge how attitudes have evolved.

A third limitation of the study lies in the variability of the academic majors of the participating students. Since the nature of the various academic disciplines vary significantly in relation to their interaction with generative AI (e.g. based on the characteristics of their course assignments and assessments), there is a possibility that students from some majors might be predisposed to more frequent use and acceptance of AI compared to students from other majors. Such differences in majors could potentially explain some of the variation observed in the data as well.

A fourth limitation of the study relates to the linguistic aspect of languages and ChatGPT. On the one hand, ChatGPT tends to be slower and less linguistically accurate in Greek compared to English. On the other hand, it was not widely known at the time that ChatGPT can be used in languages other than English. Therefore, it is likely that some of the Cypriot participants might have not been aware that ChatGPT is available in Greek, which could also explain its low rates of usage among the students in that sample.

Due to these limitations, and since this was not an experimental study, our data does not allow us to infer any causal relationships. While we have endeavored to control for confounding factors like demographics as much as possible, there could still be unmeasured heterogeneity that generates upward or downward bias. Nonetheless, the survey provides an important starting point/baseline related to ChatGPT use in the current time period and provides results that can help guide future research, curricula, assessment practices, and education policy among administrators. Moreover, these considerations highlight the need for future studies to complement our work by utilizing a more diverse and randomly selected representative sample, which would further broaden our understanding of the attitudes towards and use of generative AI in academia.

6. Conclusion

As we witness the widespread adoption of generative AI tools like ChatGPT in higher education, our study confirms the necessity for a fundamental shift in assessment practices. Our international survey across

three universities reveals a shared recognition among educators and students of the profound impact AI has on traditional assessments, particularly in essay writing and coding. Educators are in favor of adapting assessments to leverage AI's capabilities to enhance critical thinking, while students express mixed feelings, highlighting concerns over potential losses in creativity. This underscores the dual challenge we face: to harness AI's power to improve learning outcomes while also preserving and fostering students' creative and evaluative capacities. By prioritizing educational strategies that emphasize cognitive flexibility, ethical reasoning, and the nuanced interplay between technology and human values, we can equip students to navigate and thrive in an AI-driven world. As Markauskaite et al. (2022) underscore, our efforts must extend beyond technological know-how and encompass a comprehensive understanding of the complex dynamics that shape our AI-driven world.

7. Statements on open data and ethics

The study was reviewed by the institutional ethics committee at each institution (Cornell University: IRB0147447; The University of Australia: HREC 2021/800; The University of Nicosia: UREC/2023/5). Informed consent was obtained from all participants and their privacy was maintained by protecting their personal information during the research process. Respondents knew that participation was voluntary and that they could withdraw from the study at any time. One of the three ethics approvals prohibits sharing data even if de-identified, except with investigators added to the ethics protocol. Researchers interested in the data should email the corresponding author.

CRediT authorship contribution statement

René F. Kizilcec: Writing – review & editing, Writing – original draft, Visualization, Project administration, Methodology, Investigation, Formal analysis, Data curation, Conceptualization. **Elaine Huber:** Writing – review & editing, Writing – original draft, Project administration, Methodology, Investigation, Formal analysis, Data curation, Conceptualization. **Elena C. Papanastasiou:** Writing – review & editing, Methodology, Investigation, Formal analysis, Data curation. **Andrew Cram:** Writing – review & editing, Investigation, Conceptualization. **Christos A. Makridis:** Writing – review & editing, Methodology, Investigation. **Adele Smolansky:** Writing – review & editing, Validation, Investigation, Formal analysis, Data curation, Conceptualization. **Sandris Zeivots:** Writing – review & editing, Methodology, Investigation, Conceptualization. **Corina Raduescu:** Writing – review & editing, Methodology, Investigation, Conceptualization.

Declaration of competing interest

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

References

Acemoglu, D., & Restrepo, P. (2018). The race between man and machine: Implications of technology for growth, factor shares, and employment. *American Economic Review*, *108*, 1488–1542.

Anderson, L., Krathwohl, D., & Bloom, B. (2001). *A taxonomy for learning, teaching, and assessing: A revision of Bloom's taxonomy of educational objectives*. New York, NY: Longman. https://books.google.com.au/books?id=EMQlAQAAIAAJ.

Andrade, H., & Du, Y. (2019). Student perspectives on rubric-referenced assessment. *Practical Assessment, Research & Evaluation*, *10*, 3.

Bearman, M., & Luckin, R. (2020). Preparing university assessment for a world with ai: Tasks for human intelligence. In *Re-imagining university assessment in a digital world* (pp. 49–63).

Bearman, M., Nieminen, J. H., & Ajjawi, R. (2023). Designing assessment in a digital world: An organising framework. *Assessment & Evaluation in Higher Education*, *48*, 291–304.

Bennett, S., Dawson, P., Bearman, M., Molloy, E., & Boud, D. (2017). How technology shapes assessment design: Findings from a study of university teachers. *British Journal of Educational Technology*, *48*, 672–682.

Bernius, J. P., Krusche, S., & Bruegge, B. (2022). Machine learning based feedback on textual student answers in large courses. *Computers and Education: Artificial Intelligence*, *3*, Article 100081.

Bevitt, S. (2015). Assessment innovation and student experience: A new assessment challenge and call for a multi-perspective approach to assessment research. *Assessment & Evaluation in Higher Education*, *40*, 103–119.

Bloom, B. S., Engelhart, M. D., Furst, E. J., Hill, W. H., Krathwohl, D. R., et al. (1956). *Taxonomy of educational objectives: The classification of educational goals. Handbook 1: Cognitive domain*. New York: Longman.

Brynjolfsson, E., Li, D., & Raymond, L. R. (2023). Generative AI at work. NBER working paper.

Chan, C. K. Y., & Chen, S. W. (2023). Student partnership in assessment in higher education: A systematic review. *Assessment & Evaluation in Higher Education*, 1–13.

Cotton, D. R., Cotton, P. A., & Shipway, J. R. (2023). Chatting and cheating: Ensuring academic integrity in the era of chatgpt. *Innovations in Education and Teaching International*, 1–12.

Cram, A., Harris, L., Raduescu, C., Huber, E., Zeivots, S., Brodzeli, A., Wright, S., & White, A. (2022). Online assessment in Australian university business schools: A snapshot of usage and challenges. *ASCILITE Publications*, Article e22181.

Dann, R. (2014). Assessment as learning: Blurring the boundaries of assessment and learning for theory, policy and practice. *Assessment in Education: Principles, Policy & Practice*, *21*, 149–166.

Eloundou, T., Manning, S., Mishkin, P., & Rock, D. (2023). GPTs are GPTs: An Early Look at the Labor Market Impact Potential of Large Language Models. arXiv:2303.10130.

Farazouli, A., Cerratto-Pargman, T., Bolander-Laksov, K., & McGrath, C. (2023). Hello gpt! Goodbye home examination? An exploratory study of ai chatbots impact on university teachers' assessment practices. *Assessment & Evaluation in Higher Education*, 1–13.

Foster, A. D., & Rosenzweig, M. R. (2010). Microeconomics of technology adoption. *Annual Review of Economics*, *2*, 395–424.

Frank, M. R., Autor, D., Bessen, J. E., Brynjolfsson, E., Cebrian, M., Deming, D. J., Feldman, M., Groh, M., Lobo, J., Moro, E., Wang, D., Youn, H., & Rahwan, I. (2019). Toward understanding the impact of artificial intelligence on labor. *Proceedings of the National Academy of Sciences of the United States of America*, *116*, 6531–6539.

Gašević, D., Greiff, S., & Shaffer, D. W. (2022). Towards strengthening links between learning analytics and assessment: Challenges and potentials of a promising new bond. *Computers in Human Behavior*, *134*, Article 107304.

Gašević, D., Siemens, G., & Sadiq, S. (2023). Empowering learners for the age of artificial intelligence. *Computers and Education: Artificial Intelligence*, *100130*.

Gikandi, J. W., Morrow, D., & Davis, N. E. (2011). Online formative assessment in higher education: A review of the literature. *Computers & Education*, *57*, 2333–2351.

Gipps, C., & Stobart, G. (2009). Fairness in assessment. In *Educational assessment in the 21st century: Connecting theory and practice* (pp. 105–118). Springer.

Hopfenbeck, T. N., Zhang, Z., Sun, Z., Robertson, P., & McGrane, J. A. (2023). Challenges and opportunities for classroom-based formative assessment and ai: A perspective article. *Frontiers in Education*.

Huber, E., Harris, L., Wright, S., White, A., Raduescu, C., Zeivots, S., Cram, A., & Brodzeli, A. (2024). Towards a framework for designing and evaluating online assessments in business education. *Assessment & Evaluation in Higher Education*, *49*, 102–116.

Kasneci, E., Seßler, K., Küchemann, S., Bannert, M., Dementieva, D., Fischer, F., Gasser, U., Groh, G., Günnemann, S., Hüllermeier, E., et al. (2023). Chatgpt for good? On opportunities and challenges of large language models for education. *Learning and Individual Differences*, *103*, Article 102274.

Krathwohl, D. R. (2002). A revision of bloom's taxonomy: An overview. *Theory Into Practice*, *41*, 212–218.

Mao, J., Chen, B., & Liu, J. C. (2024). Generative artificial intelligence in education and its implications for assessment. *TechTrends*, *68*, 58–66.

Markauskaite, L., Marrone, R., Poquet, O., Knight, S., Martinez-Maldonado, R., Howard, S., Tondeur, J., De Laat, M., Shum, S. B., Gašević, D., et al. (2022). Rethinking the entwinement between artificial intelligence and human learning: What capabilities do learners need for a world with ai? *Computers and Education: Artificial Intelligence*, *3*, Article 100056.

Markel, J. M., Opferman, S. G., Landay, J. A., & Piech, C. (2023). Gpteach: Interactive ta training with gpt-based students. In *Proceedings of the tenth acm conference on learning@ scale* (pp. 226–236).

Mate, K., & Weidenhofer, J. (2022). Considerations and strategies for effective online assessment with a focus on the biomedical sciences. *Faseb Bioadvances*, *4*, 9.

Mislevy, R. J., Steinberg, L. S., & Almond, R. G. (2003). Focus article: On the structure of educational assessments. *Measurement: Interdisciplinary Research and Perspectives*, *1*, 3–62.

Nieminen, J. H., Bearman, M., & Ajjawi, R. (2023). Designing the digital in authentic assessment: Is it fit for purpose? *Assessment & Evaluation in Higher Education*, *48*, 529–543.

Nikolic, S., Daniel, S., Haque, R., Belkina, M., Hassan, G. M., Grundy, S., Lyden, S., & Neal., P. (2023). Chatgpt versus engineering education assessment: A multidisciplinary and multi-institutional benchmarking and analysis of this generative artificial intelligence tool to investigate assessment integrity. *European Journal of Engineering Education*, 1–56.

Papanastasiou, E. C., & Solomonidou, G. (2023). Evaluating emergency remote assessment adaptations in higher education due to covid-19: Faculty insights and challenges. *Education Sciences*, *13*, 184.

Park, J. S., O'Brien, J., Cai, C. J., Morris, M. R., Liang, P., & Bernstein, M. S. (2023). Generative agents: Interactive simulacra of human behavior. In *Proceedings of the 36th annual ACM symposium on user interface software and technology* (pp. 1–22).

Perkins, M. (2023). Academic integrity considerations of ai large language models in the post-pandemic era: Chatgpt and beyond. *Journal of University Teaching & Learning Practice*, *20*, Article 07.

Sabzalieva, E., & Valentini, A. (2023). Chatgpt and artificial intelligence in higher education: Quick start guide.

Siemens, G., Marmolejo-Ramos, F., Gabriel, F., Medeiros, K., Marrone, R., Joksimovic, S., & de Laat, M. (2022). Human and artificial cognition. *Computers and Education: Artificial Intelligence*, *3*.

Sillat, L. H., Tammets, K., & Laanpere, M. (2021). Digital competence assessment methods in higher education: A systematic literature review. *Education Sciences*, *11*, 402.

Swiecki, Z., Khosravi, H., Chen, G., Martinez-Maldonado, R., Lodge, J. M., Milligan, S., Selwyn, N., & Gašević, D. (2022). Assessment in the age of artificial intelligence. *Computers and Education: Artificial Intelligence*, *3*, Article 100075.

Tai, J., Ajjawi, R., Boud, D., Dawson, P., & Panadero, E. (2018). Developing evaluative judgement: Enabling students to make decisions about the quality of work. *Higher Education*, *76*, 467–481.

Tertiary Education Quality and Standards Agency (2024). Artificial intelligence. https://www.teqsa.gov.au/guides-resources/higher-education-good-practice-hub/artificial-intelligence.

Torrance, H. (2007). Assessment as learning? How the use of explicit learning objectives, assessment criteria and feedback in post-secondary education and training can come to dominate learning. *Assessment in Education*, *14*, 281–294.

Viberg, O., Mutimukwe, C., Hrastinski, S., Cerratto-Pargman, T., & Lilliesköld, J. (2024). Exploring teachers' (future) digital assessment practices in higher education: Instrument and model development. *British Journal of Educational Technology*.

Vygotsky, L. S., & Cole, M. (1978). *Mind in society: Development of higher psychological processes*. Harvard University Press.

Weng, J. C. (2023). Putting intellectual robots to work: Implementing generative AI tools in project management. Technical Report, NYU SPS Applied Analytics Laboratory.

Yu, H. (2023). Reflection on whether chat gpt should be banned by academia from the perspective of education and teaching. *Frontiers in Psychology*, *14*, Article 1181712.

Zambrano, A. F., Liu, X., Barany, A., Baker, R. S., Kim, J., & Nasiar, N. (2023). From ncoder to chatgpt: From automated coding to refining human coding. In *International conference on quantitative ethnography* (pp. 470–485). Springer.