Educator and Student Perspectives on the Impact of Generative AI on Assessments in Higher Education

6 authors, including:

Adele Smolansky
Cornell University
3 PUBLICATIONS   **2** CITATIONS

SEE PROFILE

Sandris Zeivots
The University of Sydney
21 PUBLICATIONS   **79** CITATIONS

SEE PROFILE

Elaine Huber
The University of Sydney
43 PUBLICATIONS   **140** CITATIONS

SEE PROFILE

René F Kizilcec
Cornell University
93 PUBLICATIONS   **4,949** CITATIONS

SEE PROFILE

Some of the authors of this publication are also working on these related projects:

Project   Self-regulated Learning in MOOCs View project

Project   Connected Learning at Scale (CLaS) View project

Educator and Student Perspectives on the Impact of Generative AI on Assessments in Higher Education

Adele Smolansky
Department of Computer Science
Cornell University
Ithaca, NY, USA
as2953@cornell.edu

Andrew Cram
Business Co-Design
The University of Sydney Business School
Sydney, NSW, Australia
andrew.cram@sydney.edu.au

Corina Raduescu
Business Information Systems
The University of Sydney Business School
Sydney, NSW, Australia
corina.raduescu@sydney.edu.au

Sandris Zeivots
Business Co-Design
The University of Sydney Business School
Sydney, NSW, Australia
sandris.zeivots@sydney.edu.au

Elaine Huber
Business Co-Design
The University of Sydney Business School
Sydney, NSW, Australia
elaine.huber@sydney.edu.au

René F. Kizilcec
Department of Information Science
Cornell University
Ithaca, NY, USA
kizilcec@cornell.edu

ABSTRACT

The sudden popularity and availability of generative AI tools, such as ChatGPT that can write compelling essays on any topic, code in various programming languages, and ace standardized tests across domains, raises questions about the sustainability of traditional assessment practices. To seize this opportunity for innovation in assessment practice, we conducted a survey to understand both the educators' and students' perspectives on the issue. We measure and compare attitudes of both stakeholders across various assessment scenarios, building on an established framework for examining the quality of online assessments along six dimensions. Responses from 389 students and 36 educators across two universities indicate moderate usage of generative AI, consensus for which types of assessments are most impacted, and concerns about academic integrity. Educators prefer adapted assessments that assume AI will be used and encourage critical thinking, but students' reaction is 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, higher-order thinking, and authentic applications.

CCS CONCEPTS

• Applied computing ~ Education

KEYWORDS

Assessment, Generative AI, ChatGPT, Educators, Students, Survey

1 INTRODUCTION

Despite the long history of research and applications of Artificial Intelligence (AI) in education, it took the sudden popularity of ChatGPT for students, educators, and university administrators around the world to be compelled to grapple with its practical implications. The introduction of GPT-3, a large language model developed by OpenAI, and the corresponding ChatGPT tool with a graphical user interface sparked global interest in the potential uses and impacts of generative AI in many domains including education. Generative AI tools can write human-like text in a conversational style, and ChatGPT has been applied in language translation, chat bots for conversations with humans, writing articles, stories, computer code, and other types of writing [1].

Generative AI tools can offer many benefits in education, such as increasing student engagement in learning tasks, providing timely feedback, aiding collaboration, and improving accessibility. For example, the ability to provide immediate and meaningful feedback through automated marking is a key benefit [2]. However, AI also raises serious concerns about the validity of assessment practices, including issues of academic integrity, especially honesty and plagiarism [3]. Researchers have highlighted the potential for

plagiarism as a key challenge with using ChatGPT for assessment in higher education [1]. Students can potentially use generative AI tools like ChatGPT to cheat on online assessments by submitting essays that are not their own work. This can challenge academics' ability to distinguish between students' own work and responses generated by such tools. It also makes it harder to assess students' level of understanding and ability to apply material. Unless educators and academic institutions adapt to this new reality, generative AI can undermine academic integrity in online assessment and the purpose of higher education to educate students, and ultimately reduce the value of a university degree [1].

Adapting current assessment practices in response to the ubiquitous availability of generative AI tools is timely but also effortful. Researchers have identified several challenges educators face in designing and implementing new online assessments [4]. 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.

While this moment in time presents a rare opportunity for innovation in current assessment practices, it is important to understand both the educators' and students' perspectives on the issue to achieve sustainable improvements. We developed and conducted a survey to compare the attitudes of both stakeholders, building on an established framework for examining the quality of university assessments. Educators and students are asked to consider several assessment scenarios and indicate their preferences to help us understand their perceptions of different adaptations and inform efforts to reform assessment practices.

2   RELATED WORK

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 [5]. Building on an extensive literature review, they identified six design criteria and four contextual factors at play (Figure 1). The design criteria 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 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 [2, 4].
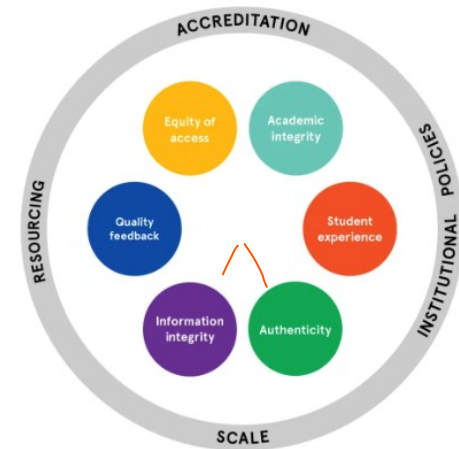
Figure 1: A framework of design criteria and contextual factors for quality online assessment [5].

In this research study, we build on a framework [5], which has thus far been tested in the Australian higher education context with instructors, by testing it as a heuristic to evaluate the impact of generative AI on assessments in new contexts. In particular, we are collecting data in a US higher education context and including students in the sample. We believe that the framework is likely to generalize across countries and we are interested in examining if it can accurately reflect students' perspectives on assessment as well. This study centers 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.

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

Dimension	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 disability/ impairment, limited access to technology, geographically dispersed)

3 METHODS

We have an ongoing survey study to investigate student and educator perspectives on the impact of generative AI tools on the design and use of online assessments. The survey instrument we developed allows us to evaluate the impact across various dimensions of online assessments [5] and evaluate similarities and differences between the student and educator perspectives.

The sample we have collected thus far consists of 389 students and 36 educators from two selective institutions of higher education, one in Australia (338 students; 26 educators) and one in the United States (51 students; 10 educators). We recruited students at different stages in their programs and in a variety of disciplines. Educators from different departments with variable assessment practices were recruited. The survey took about 15 minutes to complete. It consists mostly of scenario-based multiple-choice questions. For the U.S. sample, incentives to respond include charity donations and course credit for students; no incentives were offered for Australian respondents. Respondents who provided informed consent and answered the first set of question were retained for analysis. Responses were not required to proceed.

The survey format and questions were almost identical for educators and students since our goal is to compare their perspectives. We ask respondents about their past experience with ChatGPT and other generative AI tools and their opinions on the impact of generative AI tools on eight specific types of online assessment. We then present two sets of assessment scenarios, an essay-based assessment and a coding-based assessment:

Essay assessment prompt: 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.

Coding assessment prompt: (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.

After reading each assessment prompt, participants were asked: "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 (see Table 1). 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 respond for each dimension of the framework in Table 1 using the response options *None, Low, Medium, High, Not Sure.* Then they

are asked an open-ended question: "How do you think students might use generative AI tools to complete this assessment?"

Next, participants are given a modified version of the assessment prompt and told "consider an adapted version of the assessment prompt, given today when generative AI tools, such as ChatGPT, are widely available."

Adapted essay prompt: 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.

Adapted coding prompt: 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.

After being presented with each adapted assessment prompt, participants are asked again: "To what extent does this assessment ensure each design assessment dimension (Table 1)? This yields a clear within-participant comparison of how the adaptation affects their judgement of assessment quality. In addition, participants are asked to choose between the original prompt, adapted prompt, or indicate no preference between the two: "Assuming ChatGPT is available, which assessment do you prefer?" and "Assuming ChatGPT is available, which scenario do you think students (educators) prefer?" This not only yields a measure of their preference but also a measure of what they think the other prefers.

To analyze the data we have collected so far, we compute percentages and means of participants' ratings to observe major trends pooling across the two institutions. We present examples of common open-ended responses, prior to conducting a complete thematic analysis. This analysis focuses on examining differences between student and educator responses.

4 PRELIMINARY FINDINGS

We find that almost every student and all educators in our sample have heard of ChatGPT. However, only one in four students used it weekly or daily for coursework (29% Australia; 24% US) and for fun (25% Australia; 14% US). Educators were using it weekly or daily for professional purposes (35% Australia; 10% US), for research (15% Australia; 30% US), and for fun (31% Australia; 40% US). Besides ChatGPT, students and educators use tools like Anthropic, Bard, BingChat, ClaudeAI, DALL-E, Midjourney, and Stable Diffusion.

Students and educators rated how much different types of assessment are impacted by generative AI and there was consensus

that essays (incl. reports, literature review, case studies, research papers), computer code (incl. pseudo-code, mathematical proofs), short-answer and multiple-choice questions are very or at least moderately impacted. Assessments that require product design or creative/artistic work are also rated as moderately affected. The assignment types rated to be least impacted are presentations and discussions that are either pre-recorded or live. Educators, especially in Australia (Fig. 1), rated the impact of generative AI higher than students for all types of assignments.
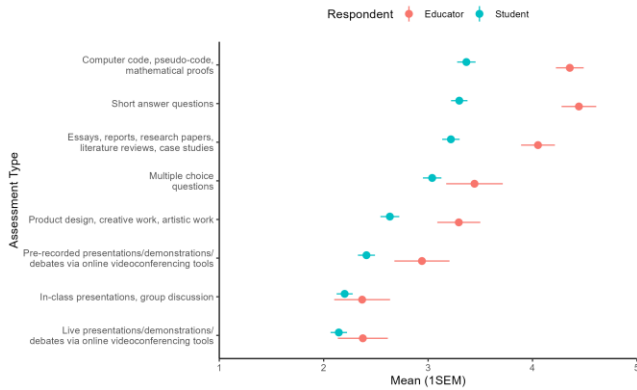
Figure 2: Australian educators' and students' ratings of how impacted different types of assessment are by generative AI. Mean ratings with standard error bars are shown.

For both specific assessment scenarios (the essay and coding prompts), students and educators reported a substantial drop in the original prompt's capacity to ensure academic integrity due to the availability of ChatGPT. The proposed adaptation to the original prompt was generally thought to restore the level of academic integrity. A similar pattern was observed for students' ratings of the assessment's authenticity. Yet ratings of the other dimensions of assessment quality shown in Table 1 remained largely stable for students and educators.

Finally, the preference ratings for the original compared to the adapted assessment prompt by each set of stakeholders pooled across countries is visualized in Figure 3. It shows that educators have a strong preference for the adapted prompt while students are less convinced by the adaptation, especially for the essay prompt.

Open-ended responses from students indicated concern about a loss in creativity in the adapted essay prompt because it gave them an essay to critique instead of asking them to write one on their own: "it kills creativity. You can't ask humans to be the secretary to machines." Students also raised concerns about the adapted coding assessment: "I would prefer to learn how to write code than learn to analyse" and "[it] replaces the teacher with ChatGPT, removing space for quality feedback." Yet students also indicated that ChatGPT is useful to get them started "[as] a prompt or inspiration but we wouldn't use it directly as a source."

Students also had forward-looking ethical concerns: "Students should be allowed to use generative AI tools like a tool to boost their productivity as long as they can be accepted in their future workplace." For the coding assessment, students noted that they are already using various tools: "before using AI tools, students have been using Google or GitHub to search for similar codes, which is also widely accepted in the workplace."
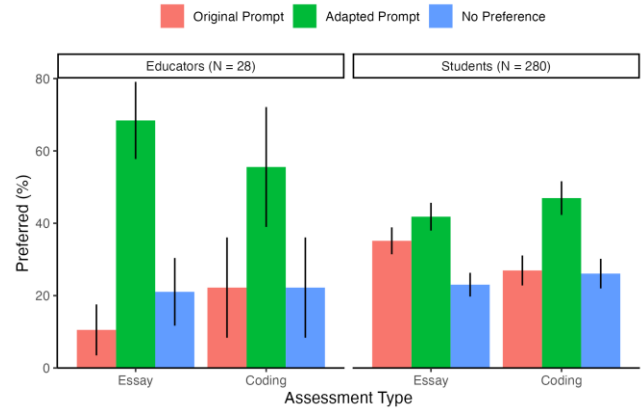
Figure 3: Educators' and students' preferences for an original or adapted assessment prompt for an essay and coding scenario. Standard error bars are shown.

5 DISCUSSION

This ongoing study uses a theoretically grounded instrument to understand educator and student attitudes about the impact of generative AI on assessments. As we collect more data in the coming months, we plan to examine variation in attitudes across contexts, backgrounds, and prior experiences. We expected to find variation between educators and students and across countries in line with different educational approaches and cultural norms. Thus far, we have observed a surprising level of agreement across stakeholders and locals. An important exception is that students are more hesitant to endorse the adapted prompt than educators for reasons including a perceived loss in creativity. 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.

There is much work to be done in preparing our students for a world in which AI tools are ubiquitous [6]. Indeed, one student in our sample wrote, "I think universities should be more interested in teaching us how to use more ways to solve problems". We concur with other researchers that there is a need to build students' capabilities in AI, but more importantly, there is a need to help them navigate the more complex interplay between technology, cognition, social interaction and values [6].

ACKNOWLEDGMENTS

REFERENCES

[1]   Debby R. E. Cotton, Peter A. Cotton, and J. Reuben Shipway. 2023. Chatting and cheating: Ensuring academic integrity in the era of ChatGPT. *Innovations in Education and Teaching International* 0, 0 (March 2023), 1–12. DOI:https://doi.org/10.1080/14703297.2023.2190148

[2]   Karen Mate and Judith Weidenhofer. 2022. Considerations and strategies for effective online assessment with a focus on the biomedical sciences. *FASEB BioAdvances* 4, 1 (2022), 9–21. DOI:https://doi.org/10.1096/fba.2021-00075

[3]   Zachari Swiecki, Hassan Khosravi, Guanliang Chen, Roberto Martinez-Maldonado, Jason M. Lodge, Sandra Milligan, Neil Selwyn, and Dragan Gašević. 2022. Assessment in the age of artificial intelligence. *Computers and Education: Artificial Intelligence* 3, (2022), 100075. DOI:https://doi.org/10.1016/j.caeai.2022.100075

[4]   Andrew Cram, Lynne Harris, Corina Raduescu, Elaine Huber, Sandris Zeivots, Andrew Brodzeli, Sue Wright, and Amanda White. 2022. Online Assessment in Australian University Business Schools: A Snapshot of Usage and Challenges. *ASCILITE Publications* (November 2022), e22181–e22181. DOI:https://doi.org/10.14742/apubs.2022.181

[5]   Elaine Huber, Lynne Harris, Sue Wright, Amanda White, Corina Raduescu, Sandris Zeivots, Andrew Cram, and Andrew Brodzeli. 2023. Towards a framework for designing and evaluating online assessments in business education. *Assessment and evaluation in higher education* (2023). DOI:https://doi.org/10.1080/02602938.2023.2183487

[6]   Lina Markauskaite, Rebecca Marrone, Oleksandra Poquet, Simon Knight, Roberto Martinez-Maldonado, Sarah Howard, Jo Tondeur, Maarten De Laat, Simon Buckingham Shum, Dragan Gašević, and George Siemens. 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, (January 2022), 100056. DOI:https://doi.org/10.1016/j.caeai.2022.100056