# Engaging and Effective Assessments in the Age of Generative AI: Student Perceptions and Recommendations

**Dr. Danushka Bandara, Fairfield University**

DANUSHKA BANDARA received the bachelor's degree in Electrical Engineering from the University of Moratuwa, Sri Lanka, in 2009. He received his master's and Ph.D. degrees in Computer Engineering and Electrical and Computer Engineering from Syracuse University, Syracuse, NY, USA, in 2013 and 2018, respectively. From 2019 to 2020, he worked as a Data Scientist at Corning Incorporated, Corning, NY, USA. Currently, he is an Assistant Professor of Computer Science and Engineering at Fairfield University, Fairfield, CT, USA. His Current research interests include Applied machine learning, Bioinformatics, Human-computer interaction, and Computational social science.

**Dr. Djedjiga Belfadel, Fairfield University**

Djedjiga Belfadel is an Associate Professor in the Electrical and Bio Engineering department at Fairfield University. She obtained her Ph.D. degree from University of Connecticut in 2015, in electrical engineering. Her interests include embedded system

# Engaging and Effective Assessments in the Age of Generative AI: Student Perceptions and Recommendations

Danushka Bandara, and Djedjiga Belfadel
Fairfield University

## Abstract

This work in progress paper discusses assessment redesign which was done for a graduate level machine learning course to address challenges of generative AI. The role of assessments in higher education is pivotal for evaluating student learning, encouraging critical thinking, and preparing learners for real-world challenges. Traditional assessment methods, such as exams and essays, have long been central to educational evaluation. However, the rise of generative artificial intelligence (GenAI) tools like ChatGPT has disrupted this landscape, offering students unprecedented capabilities to generate content, simplify concepts, and automate problem-solving. While these tools hold potential to enhance learning, they also pose significant challenges to academic integrity and the validity of traditional assessments. This study explores the redesign of assessments to address these challenges, focusing on scaffolded, multimodal, and real-time formats. Drawing from a graduate-level course in Machine Learning, the research examines student perceptions of traditional and redesigned assessments, their engagement levels, and the role of GenAI. A mixed-methods approach, combining quantitative and qualitative data, was employed to analyze feedback from 12 students as a preliminary study. Results indicate that redesigned assessments, particularly scaffolded projects, are more engaging and effective in promoting critical thinking and real-world application compared to traditional methods. Students also recognized GenAI as both a valuable learning tool and a potential risk to academic integrity. These findings contribute to the ongoing discourse on adapting higher education to the challenges and opportunities presented by GenAI, ensuring assessments remain relevant and effective in fostering meaningful learning outcomes.

## 1. Introduction

The role of assessments in higher education is fundamental to evaluating student learning, fostering critical thinking, and preparing learners for real-world challenges. Traditional assessment methods—such as homework essays, and problem sets—have long been regarded as the cornerstone of educational evaluation. However, as education evolves to meet the needs of an increasingly technology-driven world, these conventional approaches face growing challenges.

Simultaneously, the advent of generative artificial intelligence (GenAI) tools, such as ChatGPT, Claude, and Gemini, has brought transformative changes to the educational landscape. When ChatGPT was first introduced in 2023, it affected a fundamental shift in the role of the educator. These tools provide students with powerful capabilities, including generating content, simplifying complex concepts, and automating problem-solving processes. While GenAI tools have immense potential to enhance learning and teaching experiences, they also pose significant challenges, particularly in the context of assessments (Swiecki, et al., 2022). Students now have the ability to rely on AI-generated work to complete assignments and exams, raising concerns about academic integrity, the authenticity of submissions, and the validity of traditional methods in accurately measuring student learning outcomes. It is clear that, if traditional methods of assessments are not adjusted to match with this new reality, it

can seriously affect the quality of student learning and ultimately undermine the whole enterprise of higher education (Cotton et al., 2024).

In response to these challenges, educators and institutions must rethink how assessments are designed and implemented to ensure they remain relevant and effective. However, the early response of higher education institutions has been either a blanket ban on using these tools (Chan , 2023), or allowing its use as long as they are cited or acknowledged (Sharples, 2022). It is clear that this leaves educators with a lot of leeway in terms of how they handle the use of gen AI by students. This can both be an opportunity and a curse.

The integration of innovative assessment practices—such as authentic assessments, scaffolded projects, and oral defenses—provides opportunities to address the limitations of traditional methods while mitigating the influence of GenAI. Authentic assessments, for instance, focus on real-world problem-solving and critical thinking, making it more difficult for AI tools to replicate genuine student effort. Scaffolded projects encourage step-by-step development of skills and knowledge, emphasizing the learning process rather than just the final product. Documenting this process via journaling and reflection can also help in this regard. Oral defenses, on the other hand, ensure authenticity by requiring students to verbally articulate their understanding and defend their work.

Smolansky et al. (2023) lists several challenges in assessment reimagining for genAI: 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. These additional considerations also place increasing demands on faculty time and workload due to the dynamic nature of these technologies. The author's experience in assessment reimagining is in close agreement with these claims. However, the main goal of this study is to assess the student perspective and response to these changes.

This paper seeks to explore the redesign of assessments in higher education to enhance their resilience against the influence of GenAI while maintaining academic rigor and fostering meaningful learning outcomes. Drawing from the author's experience in teaching undergraduate and graduate-level courses in Machine Learning and Data Mining, this study offers insights into the vulnerabilities of traditional assessment practices and the effectiveness of redesigned approaches. The results from a preliminary study conducted during the Fall semester of 2024 are presented. By presenting objective and subjective quantitative comparisons of student performance and attitudes, this research aims to Explore student perceptions of traditional and redesigned assessments, their engagement levels, and the role of AI tools.

The research questions addressed in this study are,

1. How effective are traditional and redesigned assessments in supporting learning outcomes?
2. What are the challenges and benefits of redesigned assessments?
3. How do students perceive the role of generative AI in assessments?

Additionally, the paper presents best practices based on the results of the study and existing literature for incorporating GenAI into the learning process as a supportive tool rather than a means of bypassing genuine effort. It emphasizes the importance of balancing the benefits of AI-driven technologies with the need to cultivate core skills and knowledge in students. Ultimately, this work contributes to the ongoing discourse on how higher education can adapt to the challenges and opportunities presented by GenAI, ensuring that assessments continue to play a meaningful role in shaping the next generation of learners and professionals.

## 3. Methodology

During the semester, the students were assigned several assessments in the following three formats (Scaffolded, Multimodal and Realtime). Examples of redesigned assessments are given below,

Scaffolded assessment: A step by step assessment which builds up incrementally

*Question1: Explain the difference between classification and regression problems in machine learning. Provide one example for each type of problem.*
*Question 2: Below is a sample dataset showing customer information (e.g., age, income, and whether they purchased a product).*
*Identify the target variable and explain why you chose it.*
*Suggest three features from the dataset that might influence the target variable and explain your reasoning.*
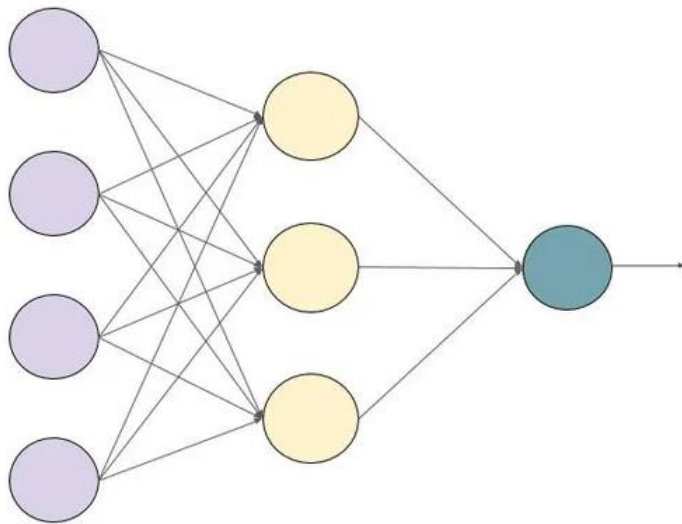*Question 3: Preprocessing the Data*
*Identify any missing values or outliers in the dataset.*
*Write pseudocode to handle missing data and scale numerical features.*

Multimodal assessment: Using multiple forms of input as parts of the assessment

*Example Question: Below is a diagram of a simple feedforward neural network. Annotate the diagram by identifying the input layer, hidden layers, and output layer. Then, explain the role of activation functions in this architecture.*

Realtime assessment: Completed within class during a limited period of time

*Example Question: You are provided with a dataset in a Jupyter notebook during class. Perform the following in 20 minutes:*
- *Preprocess the dataset.*
- *Train a decision tree classifier.*
- *Evaluate its performance using a confusion matrix.*

**Student feedback survey**

An anonymous survey was designed and administered to the students to gather both quantitative and qualitative feedback from students regarding their familiarity with and attitudes toward generative AI tools, their preferences between traditional and redesigned assessment methods, and the impact of these tools on their learning experiences. To achieve this, the survey incorporated a combination of Likert-scale questions and open-ended responses, allowing for a comprehensive understanding of student perspectives. To minimize potential bias, the survey was conducted anonymously and administered after final grades had been submitted. Given the limited class size, this study represents a preliminary, small-scale investigation designed to provide early insights into student perceptions and engagement with redesigned assessments in the context of GenAI. The participants in this study were enrolled in a graduate-level course, with a total of 12 students. Among them, one was an undergraduate student, one was pursuing an M.S. in data science, and the remaining ten were M.S. in software engineering students. For data analysis, a mixed-methods approach was employed, utilizing rating scale analysis for quantitative data and open-ended response analysis for qualitative insights. This dual approach ensured a thorough examination of the students' experiences and perceptions.

## 4. Results

### 4.1. Familiarity and Use of AI Tools

The mean familiarity score with generative AI tools is 4.45 on a 5-point scale, with a standard deviation of 0.82. This indicates that most students are highly familiar with these tools (in their opinion). The

relatively low standard deviation suggests that there is little variability in responses, meaning that students' familiarity levels are consistently high across the group. Overall, these results highlight the widespread exposure and comfort that students have with generative AI tools in their academic or personal contexts.

**4.2. Effectiveness of Traditional and Redesigned Assessments**

| Assessment Type | Mean Rating | Standard Deviation |
|---|---|---|
| Traditional Assessments | 3.45 | 1.13 |
| Real-time Assessments | 4.00 | 0.89 |
| Scaffolded Projects | 4.18 | 0.87 |
| Multimodal Assignments | 4.09 | 1.04 |

Table 1: Summary of student responses on the effectiveness of Traditional and Redesigned Assessments

The quantitative analysis reveals that students have mixed opinions on the effectiveness of traditional assessments, with a mean rating of 3.45 and a standard deviation of 1.13. The moderate variability in responses suggests that while some students find traditional assessments effective, others are less convinced. In contrast, engagement levels for redesigned assessments are consistently higher. Real-time assessments have a mean engagement rating of 4.00 ($\sigma = 0.89$), scaffolded projects score slightly higher at 4.18 ($\sigma = 0.87$), and multimodal assignments follow closely with a mean of 4.09 ($\sigma = 1.04$). These results indicate that students find redesigned assessments—particularly scaffolded and multimodal ones—more engaging than traditional methods, with relatively low variability in their positive responses.

Based on the qualitative responses, scaffolded projects were widely appreciated by students, who found them engaging and effective for learning. The structured guidance and iterative feedback provided in these assignments helped students build confidence and understanding, making complex tasks more manageable. In contrast, Multimodal Assignments elicited mixed responses. While some students enjoyed the creativity and variety offered by tasks involving images, videos, or other media, others preferred more traditional formats, possibly due to familiarity or ease of use. Real-time assessments, such as quizzes or activities during lectures, were perceived as less engaging by some students, likely because of the pressure to respond immediately and the lack of time for reflection (unless specifically built into the assessment). Overall, scaffolded projects emerged as the most consistently favored format, while multimodal and real-time assessments showed potential but may require further refinement to meet diverse student preferences.

Students perceive that redesigned assessments significantly improve their critical thinking skills, with a mean rating of 4.27 and a standard deviation of 0.90. Additionally, they believe these assessments enhance their ability to apply course concepts to real-world situations, as reflected by a mean rating of 4.18 and a standard deviation of 0.87. Overall, these findings highlight the value of redesigned assessments in increased perception of deeper learning and practical application of knowledge.

**4.3. Challenges in Redesigned Assessments**

Students face several challenges with redesigned assessments, as highlighted in their feedback. One significant issue is thinking outside the box, particularly for graduate students who expressed difficulty

in approaching tasks creatively or divergently when they lacked sufficient experience with the topic. Another challenge is the time and effort required for these assessments. Students acknowledged that redesigned assignments often demand more preparation due to the change in format and the difference in the approach required. Additionally, some students noted that the depth of understanding required for these assessments can feel overwhelming. One respondent mentioned that assessments sometimes seem too advanced or "deep," making it hard to connect them to the material covered in class. These challenges highlight the need for clear instructions, adequate support, and a balance between complexity and accessibility to ensure that redesigned assessments are both effective and manageable for students.

### 4.4. Role of Generative AI in Assessments

Students recognize generative AI tools as valuable learning tools, often using them to better understand challenging concepts. However, they also acknowledge a potential downside: AI can sometimes provide quick answers without fostering deeper understanding, which may hinder long-term learning. While many students feel confident in their ability to use AI tools ethically, some express concerns about over-reliance on AI for completing assignments, which could compromise academic integrity. Opinions on how well AI use is managed in assessments are mixed. Some students believe that redesigned assessments—such as scaffolded projects or oral defenses—effectively minimize the risk of AI misuse by requiring critical thinking and authentic engagement. Others, however, suggest that further improvements are needed to ensure assessments remain resistant to AI-driven shortcuts. These insights highlight the dual role of AI as both a helpful resource and a potential challenge, emphasizing the need for thoughtful assessment design and clear guidelines on ethical AI use.

### 4.5 Instructor load as a result of assessment redesign

From the author's experience during the teaching of the machine learning course, scaffolded and multimodal assessments take more instructor time due to the complexity involved. This includes time that the instructor has to spend ensuring that the instructions are clear and straightforward, yet the assignment cannot be directly answered by GenAI. Real time assessments take the least amount of preparation as they mimic the traditional assessments.

# 5. Addressing the Research Questions

For RQ1: Students had mixed feelings about the effectiveess of traditional assessment. While redesigned assessments were consistently rated higher in effectiveness.

For RQ2: While redesigned assessments offer significant benefits such as higher engagement and effectiveness, they also present challenges, such as increased time and effort for both students and instructors.

For RQ3: Students perceive generative AI as a double-edged sword—valuable for learning but potentially harmful to academic integrity if not managed carefully.

# 6. Discussion

Educators can enhance student learning by adopting scaffolded assessments that break complex tasks into manageable steps, providing clear instructions and iterative feedback. Multimodal assignments

should balance creativity with accessibility, offering resources for unfamiliar tools. To address challenges like creative thinking and time management, educators should align assessment depth with course content and support students with examples and templates.

What was made abundantly clear from student feedback is that, although they understand the challenge of generative AI for assessments, they also do not want to spend more time and effort on assessments (thus the main motivation for using shortcuts such as Gen AI). Clear communication of learning goals and the value of redesigned assessments fosters a growth mindset and ensures students see the benefits of these innovative approaches.

## 7. Recommendations

In addition to assessment redesign detailed in this study, educators can also use other best practices to guard against generative AI in assessment. These are included here even though it was not a part of this particular study.

Flipped classrooms: with most of the assessments done during in person class time and homework is for watching/reading the class material. Similar to real time assessments, this ensures students are present in class and actively engaged rather than using generative AI to complete the assessments. This method also has the added advantage that students can be given time to process the material ahead of time before completing the assessments (Huang et al., 2023).

Project-Based Scaffolded Assessments: Assign long-term, multi-step projects that require students to apply course concepts to real-world problems (Jones, 2019). Projects often involve unique, context-specific tasks that are difficult for AI to replicate. They also encourage collaboration, research, and creativity. The author has previously used group based scaffolded assignments for coding assignments before, and shuffled the code between groups after each step to ensure students get to work on debugging unfamiliar code. This would be a good practice for working with generative AI as well because it encourages critical thinking and adaptation. Both skills which are important in the age of generative AI.

Oral Assessments and Presentations : Replace written submissions with oral presentations, interviews, or defenses of their work. Oral assessments require students to demonstrate their understanding in real-time (Nallaya et al., 2024), making it difficult to rely on AI-generated content. However, the drawback of this is that instructor load increases significantly. This can be a solution for small classes.

Submit assessments to a GenAI software ahead of assigning to check if it can be directly answered. If that is the case, it should not be a take home problem.

## 8. Limitations and Future Research

The study has several limitations, including a small sample size of 12 students, which restricts the generalizability of the findings, and a reliance on self-reported data, which may introduce biases such as wanting to satisfy the instructor or inaccurate recall. The study was conducted on smaller problem

sets rather than semester-long projects. Additionally, the participant pool was predominantly graduate students in software engineering and data science, limiting insights into how undergraduate students or those in other disciplines might experience redesigned assessments. The short-term focus of the study also means that long-term impacts on learning and skill retention remain unexplored. Additionally, studies could examine the impact of redesigned assessments on instructor workload, student engagement, and equity and accessibility, ensuring that innovative assessment practices benefit all learners.

# References

Anthropic. (2024). Claude  [Large language model]. https://www.anthropic.com/

Google. (2024). *Gemini* [Large language model]. https://gemini.google.com/

Huang, A. Y. Q., Lu, O. H. T., & Yang, S. J. H. (2023). Effects of artificial intelligence–enabled personalized recommendations on learners' learning engagement, motivation, and outcomes in a flipped classroom. Computers & Education, 194, 104684. https://doi.org/10.1016/j.compedu.2022.104684

Jones, B. (2019). Good practice: Scaffolded, collaborative project-based learning. Journal of the European Honors Council, 3(1), 1-16.

Nallaya, S., Gentili, S., Weeks, S., & Baldock, K. (2024). The validity, reliability, academic integrity and integration of oral assessments in higher education: A systematic review. Issues in Educational Research, 34(2), 629–646. https://search.informit.org/doi/10.3316/informit.T2024121800014001121097284

OpenAI. (2024). ChatGPT [Large language model]. https://chat.openai.com/

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, 100075. https://doi.org/10.1016/j.caeai.2022.100075

Cotton, D. R., Cotton, P. A., & Shipway, J. R. (2024). Chatting and cheating: Ensuring academic integrity in the era of ChatGPT. Innovations in education and teaching international, 61(2), 228-239.

Chan, C. K. Y. (2023). A comprehensive AI policy education framework for university teaching and learning. International journal of educational technology in higher education, 20(1), 38.

Sharples, M. (2022). New AI tools that can write student essays require educators to rethink teaching and assessment. Impact of Social Sciences Blog. https://blogs.lse.ac.uk/impactofsocialsciences/2022/05/17/new-ai-tools-that-can-write-student-essays-require-educators-to-rethink-teaching-and-assessment/