Kashihara, A. et al. (Eds.) (2024). Proceedings of the 32nd International Conference on Computers in Education. Asia-Pacific Society for Computers in Education

# Competency-Based Assessment in the Era of Generative Artificial Intelligence: Perspectives of Selected STEM Educators

**Friday Joseph AGBO[a*], Heather Kitada SMALLEY[a] & Kathryn NYMAN[c]**
[a]*School of Computing and Information Sciences, Willamette University, USA*
[b]*College of Arts and Sciences, Willamette University, USA*
*fjagbo@willamette.edu

**Abstract:** Generative Artificial Intelligence (GenAI) has come to stay, and educators are exploring its usage in diverse contexts. One pertinent question begging for an answer is how educators integrating GenAI tools can equitably assess students' learning outcomes. This study explores the mixed-method approach, consisting of a rapid literature review and an analysis of experts' perspectives to address this question. We analyze data from the Scopus and Web of Science databases from the rapid review to understand how the use of GenAI is penetrating the STEM field. On the other hand, the thematic analysis of data generated from a ten-week-long group learning circle discussion among STEM professors regarding assessment in the era of the GenAI was used to gain understanding of educators' perspectives regarding how students' learning could be assessed in a classroom where GenAI tools are used. Our findings provide insights regarding how, where, and when to integrate GenAI in STEM classes and potential assessment strategies that could foster trust and transparency between educators and students. This study contributes to the growing body of literature on GenAI in STEM education. It offers implications from the perspective of contextual adoption of assessment strategy in the era of GenAI rather than the traditional approach of one-size-fits-all.

**Keywords:** Generative Pre-trained Transformer, GenAI, Large language model, Artificial Intelligence, ChatGPT, Assessment, Self-regulated learning

## 1. Introduction

Students' learning outcomes, learning process, and satisfaction are difficult to measure in any learning setting. While ongoing discussions exist on measuring students' performance and making the process more inclusive, the emergence of advanced technologies such as Generative Artificial Intelligence (GenAI) makes it more daunting to assess students' learning outcomes. GenAI tools such as ChatGPT and Copilot are gaining massive ground and transforming the learning process, making it more difficult for educators to maintain traditional teaching and assessment strategies.

Nowadays, abundant evidence showcases the integration of GenAI tools in colleges to teach STEM (Science, Technology, Engineering, and Mathematics) courses, suggesting that students can, for example, use these tools to solve coding assignments without understanding the fundamentals of problem-solving to build programming skills. Consequently, measuring students' competency in this era may go beyond the traditional method – where programming assignment evaluation, for example, is based on rubrics on correctness and error-free codes – since GenAI tools can generate correct and workable codes for an assignment. Therefore, a more dynamic and equitable assessment approach is needed, where students who do not use GenAI tools on an assignment may not suffer from low evaluations compared to those who use GenAI. This study used a mixed research method, including a rapid review of literature and viewpoints of STEM educators, to address the research question formulated. We believe that this study is one of the multiple ways educators can develop contextual strategies for assessing students in STEM classrooms that integrate GenAI tools and address the gap brought about by the advancement of these new technologies.

## 2. Background

GenAI is a term used to describe a powerful tool built with complex models to generate high-quality, human-like materials - text, images, audio, and even videos. Although AI models that mimic human intelligence have been around for decades, a recent advancement of this technology was the introduction of 'Transformers' that use Natural Language Processing (NPL) techniques (Roman, 2023). These technologies, such as GPT (Generative Pre-process Transformer) models developed by OpenAI, are capable of not just generating educational content but also performing personalized assessments of student learning (Moorhouse et al., 2023; Radovi ́c et al., 2024). However, it requires both educators and learners to be knowledgeable and familiar with these tools to make meaningful use of them. While GenAI offers abundant resources to facilitate learning, there are concerns about how it should be used in the classroom (Kohnke, 2024; Law, 2024).

Automating assessment with the integration of GenAI tools is a promising area where technology can improve educators' efficiency and reduce workload (Pinargote et al., 2024; Duane, 2024). However, studies have shown that this design can be problematic as GenAI, such as ChatGPT, could not solve some assignments and exams in certain disciplines (Chau et al., 2024; Feng et al., 2024). This inability to address specific domain problems poses a challenge, and we cannot rely on GenAI for automated solutions in all domains. Using GenAI to assess learners limits human touch, empathy, and other perspectives that machines may not perceive. This could create barriers and a lack of trust between educators and students Luo (2024).

In a recent study, Ogunleye et al. (2024) revealed that GenAI tools exhibit subject knowledge, have problem-solving analytical skills, and possess critical thinking abilities but can also deprive learners of gaining these skills if used unethically. Another study conducted by Finnie-Ansley and his colleagues reported that students' over-reliance on GenAI to generate codes for programming assignments was evident, posing concerns for educators. In addition, the authors found that the GenAI tool used in the study could not solve some of the programming problems correctly (Finnie-Ansley et al., 2022).

While this study focuses on the STEM field, studies on using GenAI for non-STEM education exist. For example, a systematic review study by Wu and Zhang (2024) revealed several studies in that domain and suggested important concerns for educators and students using GenAI tools, including 1) ethical and moral considerations, 2) GenAI should augment rather than replace human intelligence, 3) GenAI should be used as an instructional tool rather than a fully automated system 4) GenAI could improve academic *assessment* and self-*assessment* methods, and 5) review the quality of results generated by GenAI tools.

In summary, findings from these studies suggest that educators must find ways to innovatively introduce GenAI into their classrooms and not rely on them for assessment, as a one-size-fits-all approach may not be a suitable way to benefit from this technology. Regarding assessing students' learning progress while they use GenAI tools in their STEM classes, there is no clear evidence from the literature that suggests a pointer to an ideal approach. Thus, this study addresses this gap by engaging STEM educators (key stakeholders in students' education) to understand their perspectives.

***Research question:***
*Should the use of GenAI in STEM classes be deliberate? If so, how do we equitably assess students' learning?*

In this study, we refer to equitable learning assessment as a process that evaluates all students' learning progress and achievement in a way that is fair, inclusive, and responsive to their diverse needs, backgrounds, and abilities. The first part of this question was answered through a rapid literature review, while the second part was addressed through STEM educators' perspective analysis.

## 3. Methodology

This study employed a mixed-method approach consisting of a rapid review of the literature and an analysis of educators' perspectives to address the research question. According to Khangura et al. (2012), cited by Hamel et al. (2021), "a rapid review is a type of knowledge synthesis in which components of the systematic review process are simplified or omitted to produce information in a short period of time." While the rapid review aims to address the question of whether GenAI could be allowed in STEM classrooms, experts' perspectives address how to develop strategies to assess students' learning processes and outcomes in settings where GenAI tools are used to facilitate learning. In this section, we outline the process followed by these two methods.

### *3.1. Rapid Review of Literature*

We conducted a search on two main databases (Scopus and Web of Science) to collect relevant articles that discussed the theme of this study. The authors selected these two databases because they are well known for indexing STEM education papers published by reputable journals (Sanusi et al., 2021) and should contain relevant articles that are suitable for a rapid review (Hamel et al., 2021). Our search strings consisted of "GenAI," "ChatGPT," "Large Language Model," "Generative Pre-trained Transformer," "assessment," "grading," "evaluation," and "learning outcomes." The search was conducted on August 16, 2024. Our query constructs are customized using logic operators ('AND' and 'OR') to fit each database, and the results obtained from them are presented in Table 1.

Table 1. *Data search constructs and the results obtained*

| **Database** | **Query strings and constructs** | **Query result** | **Relevant articles** |
|---|---|---|---|
| Scopus | TITLE-ABS-KEY (("GenAI" OR "ChatGPT" OR "Large Language Model" OR "Generative Pre-trained Transformer") AND ("Science, Technology, Engineering, and Mathematics" OR "STEM")) AND (("assessment" OR "grading" OR "evaluation" OR "learning outcomes" )) | 53 | 10 |
| Web of Science | ("GenAI" OR "ChatGPT" OR "Large Language Model" OR "Generative Pre-trained Transformer") (Title) and ("assessment" OR "grading" OR "evaluation" OR "learning outcomes") (Topic) AND ("Science, Technology, Engineering, and Mathematics" OR "STEM") (All Field) | 5 | 1 |

Our selection criteria included peer-reviewed articles that deal with the theme of this study as presented in the research question, are written in English, focus on teaching and learning, and discuss assessment strategy in one form or another. While the research question in this study is not entirely focused on using GenAI tools for automatic learning assessment, we decided to review articles that discuss automated systems for assessment using GenAI tools in STEM education settings to address the first part of the question. Meanwhile, the second part of the question deals with students' assessment of learning outcomes in STEM classroom settings where GenAI tools are used to facilitate learning. Consequently, this study included articles that showcased using GenAI to automate assessment in its analysis.

The first author read through the abstract and research questions of the articles to determine their relevance and synthesized the data following the process demonstrated in similar studies (Braun and Clarke, 2023; Agbo, 2024) and finally applied a content analysis procedure (Sanusi et al., 2024) to analyze selected articles that met the defined criteria.

*3.2. STEM educators' perspective analysis*

Three STEM professors (the authors of this paper) organized a weekly discussion meeting under the umbrella of a 'learning circle' summer program (see Keane for an introduction to learning circles), which lasted for ten weeks. The meeting provides an opportunity to engage, brainstorm, learn, and explore ways to foster understanding and use of GenAI in our classes and to provide equitable assessment for the students. Each meeting lasted for an hour. While the first author coordinated the meeting, any of the professors could lead the topic of discussion. We dedicated time to discuss sample homework or curriculum designed by each professor to teach in their respective classes. Table 2 presents the characteristics of the professors and their different STEM disciplines. Our data collection consists of meeting notes, STEM course syllabi used as case study discussions, links to existing students' past submissions, and live demos with GenAI tools such as ChatGPT and Gemini. We analyze these data by thematically coding and synthesizing them. The findings are discussed in the next section.

Table 2. *Learning circle members and characteristics*

| **Academic Position** | **STEM Discipline** | **Course area and case study** | **Year/level (credit hour)** |
|---|---|---|---|
| Professor and Chair of the Mathematics Department | Mathematics | Interdisciplinary Studies: Escape room project | Year 1 (4 credits) |
| Assistant Professor of Statistics and Data Science | Statistics and Data Science | Data Visualization and Presentation | Masters (4 credits) |
| Assistant Professor of Computer Science | Computer Science | Human-Computer Interactions | Year 2-4 (4 credits) |

## 4. Results and Discussion

This section presents and discusses findings based on the analysis of data obtained from the two methodologies presented above to address the research question raised in this study. The first part of the presentation deals with the findings from the rapid review, while the second part presents STEM experts' perspective analysis.

*4.1. Findings based on the literature data*

This section presents a synthesized analysis of the 11 articles selected for the rapid review. The scope of the articles ranges from multidisciplinary to specific domains such as computing education, mathematics education, and data science and machine learning education. The majority of the studies were conducted in higher education settings, although K-12 and graduate settings were also represented. As shown in Table 3, the authors have investigated varied aspects of the integration of GenAI in STEM education and reported findings discussed in this section.

Table 3. *Characteristics of articles included in the rapid review*

| **Article** | **Pub. Year** | **Discipline or scope** | **Educational level** | **Study focus and findings** |
|---|---|---|---|---|
| Lee et al., (2024). | 2024 | Science Education | K-12 | This study explored the use of ChatGPT-3 and ChatGPT-4 combined with Chain-of-Thought (CoT) in developing automatic scoring. The result shows that this integration can produce |

| Article | Pub. Year | Discipline or scope | Educational level | Study focus and findings |
|---|---|---|---|---|
| | | | | explainable and interpretable automatic scoring with CoT enhancing its accuracy. |
| Algahtani (2024) | 2024 | Scientific writing | K-12 | This paper examined the perspective of STEM educators in exploring GenAI for automation in the grading process (test planning, item generation, test administration, and scoring), and predictive analytics. Findings from this study provided areas of concerns including emphasis for transparent and ethical AI algorithms, personalized and adaptive assessment, and the importance of human judgment in AI-powered STEM education. |
| Neve and Pawelczak (2024) | 2024 | Computing education | Higher education | This study explains educator's experience of working with GenAI tools for the creation of programming lab assignments for graduate and undergraduate classes. The study reported the possibilities and risks of working with this technology. |
| Zheng et al. (2023) | 2023 | Science education | Higher education | This study revealed the weakness of GenAI tools in assessing Multiple Choice Questions (MCQs) due to their 'selection bias' where they are vulnerable to option position changes. The authors propose a 'label-free, inference-time debiasing method,' called 'PriDe' to mitigate this vulnerability tendency exhibited by LLMs in dealing with MCQs |
| Eva et al. (2024) | 2024 | Multidisciplinary | Higher education | In this study, the authors investigated the perspectives of educational stakeholders regarding innovations in learning, student engagement, and assessment concerns in the GenAI era. Analyzing the qualitative data obtained through interviews suggests the viability of using GenAI for learner-centered education; however, its ethical usage is part of the stakeholders' concerns. |
| Luo (2024) | 2024 | Multidisciplinary | Higher education | This study examined the concept of 'erosion of trust' between students and educators in the era of GenAI. Using concept mapping activities interviews, the author investigated how students navigate ‘trust-building’ with teachers in this AI-mediated assessment era. Finding shows that there is a transparency issue in teacher-student trust-building which makes students feel unsafe to freely explore GenAI use. |
| Wolfer (2023). | 2023 | Computing education | Higher education | In this study, the author provides a qualitative assessment sample code generated by ChatGPT to showcase its use for programming education and reports implications for computing pedagogy. |
| Drori et al. (2023) | 2023 | Data science and machine learning education | Higher education | This study explored the use of GenAI to generate machine learning final exams and evaluated students' perception of comparing GenAI questions with human-generated questions. Finding based on student surveys shows that machine-generated questions are indistinguishable from human-generated questions and are suitable for final exams. |
| Martínez-Téllez et al., (2023) | 2023 | Mathematics education | Higher education | This study reported on the use of Bing Chat to foster flipped learning using collaborative activity in a Mathematics course. The study investigated the impact of the use of this tool on student's development of critical thinking skills and found this intervention promising. |
| Tan et al. (2023) | 2023 | Multidisciplinary | Higher education | These authors developed a formative assessment tool utilizing NLP and LLM to generate assessment questions and evaluate student answers. The study reported that this tool is capable of providing timely feedback to learners, offering scalable and personalized formative assessment experiences in STEM |

| Article | Pub. Year | Discipline or scope | Educational level | Study focus and findings |
|---|---|---|---|---|
| | | | | education |
| Koltovskaia et al., (2024) | 2024 | Multidisciplinary | Graduate education | This study reported the experiment conducted to examine how Iranian graduate students from STEM fields engaged with ChatGPT to revise their academic research proposals and investigated behavioral, cognitive, and affective aspects of their engagement. Findings show that participants relied on prompt engineering to engage ChatGPT, comprehended the feedback, but occasionally expressed doubts regarding its accuracy, and were satisfied with the outcomes. |

Regarding how assessment is conducted in studies that utilize GenAI, findings show that many of the exploratory studies presented in Table 3 establish an integration of GenAI tools into an assessment system (providing automatic assessment) and highlight the potential benefits of GenAI in STEM education. For example, educators with technical know-how about the integration of GenAI with other third-party or customized tools leverage its potential in automating their grading and assessments to reduce workload and manage time efficiently (Tan et al., 2023). Having the technical skills must be complemented with the ability to design the automation using equity and justice lens. As reported by many studies (Algahtani, 2024; Eva et al., 2024; Neve & Pawelczak, 2024; Zheng et al., 2023), equitable use of GenAI tools in assessment is a critical issue that must be attended to, as removing the important aspect of human judgment in GenAI-powered STEM education can lead to unintended consequences. For example, in (Algahtani, 2024) the authors express ethical concerns about the automation of assessment using GenAI. In another study, Zheng et al. (2023) found a 'selection bias' issue with the automation of MCQ assessment, which calls for concern. Thus, rather than implementing GenAI-powered assessment tools in isolation, they should complement the traditional assessment strategy with human supervision to offer equitable scoring judgment.

One way in which GenAI can support equity is in leveling the playing field for students in some situations. Students across education settings can utilize the abundant resources made available through GenAI to foster their learning experiences (Martínez-Téllez et al., 2023). As demonstrated by one of the studies in this review (Koltovskaia et al., 2024), students can use GenAI tools to refine their academic research proposal by finding suitable research questions or refining their study scope.

One of the studies reviewed investigated the important topic of 'trust-building' between teachers and students in GenAI usage (Luo (2024)). Interestingly, the author found that there is a transparency issue in teacher-student trust-building that makes students feel unsafe exploring GenAI use freely. In other words, students do not feel safe disclosing the usage of GenAI tools to the teacher since they are unsure about the transparency in the assessment process.

*4.2. Findings based on educators' perspective data*

This section presents the thematic analysis of the data collected during educator's ten-weeks long discussions. From the data coding, the following themes emerged: 'assessment type', 'scope of how, where and when to introduce GenAI', 'concerns', 'suggestions for developing usage guidelines', and 'future research area'. Table 4 presents the synthesized classification of the codes mapped with each theme.

Table 4**.** *Thematic code mapping of STEM educators' 'learning circle' discussion*

| **Assessment Type** | **Scope: How, where, and when to introduce GenAI** | **Concerns** | **Suggestions for developing guidelines** | **Future research area** |
|---|---|---|---|---|
| - Self-reflection approach<br>- Oral exams/interviews<br>- Inquiry-based approach<br>- Documentation and commenting (for coding-related exams)<br>- Regular reflection poll<br>- Peer reviewing | - Clear instruction within the syllabus for use cases<br>- defined as part of learning outcomes<br>- Automation of assessment<br>- Brainstorming and ideation exercise<br>- Proof of concepts and prototyping<br>- Group and community learning<br>- Providing personalized feedback<br>- Other fun aspects outside the learning objectives of a project that help engage students. | - Discerning between good versus bad generated information<br>- Environmental impact (energy consumption)<br>- Lack of human touch in the learning process<br>- GPT errors<br>- students circumventing learning goals | - Why and how should students use GenAI tools?<br>- Is it about "what are you learning" or how are you learning?<br>- Scaffolding purpose? | - Multidisciplinary study of faculty and students' perception on the use of GenAI tools<br>- Experimenting with different AI treatments in multiple sections of a course |

From experts' viewpoints, the educators' discussion analysis revealed important findings that could guide how we think, frame, and integrate GenAI into STEM classes. Noticeably, their perspectives are reflective of experiences they have gathered from teaching intro and advanced STEM courses, particularly within the mathematical sciences. The educators think that GenAI tools can be explored in STEM classes for brainstorming and ideation exercises, designing proof of concepts and prototyping, group and community learning, providing personalized feedback, and other engaging aspects outside the learning objectives. They, however, warned that educators should explicitly provide clear instructions in the syllabus regarding approved uses of GenAI and conversely where GenAI use would circumvent the student learning objectives. To make the usage instruction clearer and to facilitate equitable assessment, these experts opined that the STEM instructor should develop a guideline for using GenAI tools, suggesting some thought-provoking questions: "Why and how should students use GenAI tools?"; "Is it about 'what are you learning' or 'how are you learning'"?; "How can GenAI be used for scaffolding purposes"? All these questions could help in the formulation of usage guidelines.

Furthermore, in considering assessment techniques that would be less circumvented by GenAI tools, STEM educators suggest the following strategies: self-reflection, oral exams/interviews, inquiry-based group explorations, documentation and commenting (for coding-related exams), regular reflection polls, and peer reviewing. For example, in the case of coding exercises, students could be asked to annotate their code, explaining what portions of code does, or explaining any approaches that deviate from those discussed in class. The adoption of these suggested strategies must take cognizance of contextual factors that could impact the overall outcome, such as the goal and learning objective of the specific course, the level of students, etc., which are connected to the theme of scope – 'how, where, and when to introduce GenAI.' As demonstrated by Koltovskaia et al., (2024), students can reflect and judge their learning outcome from engaging with GenAI tools if they are guided through expectations defined by the educator.

Similarly, the suggestions from STEM educators align with recent findings in the literature. For example, Xia et al. (2024) show that in the context of using GenAI for teaching and learning, self-assessment is a strong strategy used to support students' learning outcomes, thereby allowing them to learn from the mistakes and errors of a learning task that GenAI suggested. These authors also highlighted that the learning tasks benefited student learning but did not involve critical thinking as expressed elsewhere (Ogunleye et al., 2024).

For the instructors, huge benefits abound by using GenAI to reduce their workload of developing teaching and instructional material. Alongside this comes the extra burden of acquiring new skills and strategies for assessment of students' learning. Thus, teachers must be prepared to learn about new assessment strategies to be able to cope. As suggested by the experts, instructors should learn how to discern between good and bad-generated information in order to provide meaningful feedback. Experts are concerned about the lack of human touch in the learning process when GenAI is deployed by instructors to facilitate some critical learning objectives or assessments. As suggested in other studies, educators using GenAI in their classrooms should focus on critical thinking, problem-solving, and creativity in designing assessment strategies (Chan, 2023; Dai et al., 2023). Moreover, the issue of hallucination and potential errors in materials generated by GenAI tools remains a concern for the experts. The existence of such errors can lead to productive classroom conversations which acknowledge that for someone with a surface-level understanding of a topic, the information composed by GenAI might sound sufficient; however, when one has more expertise, it can be clear that GenAI is being too broad or even not accurate. Consequently, inappropriate use of GenAI by new learners is often easy to spot, as the output does not align with the materials or pedagogies discussed in class.

## 5. Case Studies

The role that GenAI plays within a course depends on each course's core learning objectives. We present a case study of two graduate and undergraduate- level courses within Data Science and Computer Science programs to illustrate a diverse array of how GenAI can be used to enhance student experience, while still emphasizing the learning process.

First, we present a course in Human Computer Interaction, taught at both the undergraduate and graduate levels. These are project-oriented courses in which students design and implement interfaces through a user-centered design approach. The interfaces are prototypes of web or mobile applications. However, the learning objectives of this class focus on the process of problem identification, user-centered design and interactive interface development rather than the execution by means of computer programming. Students are asked to create “low-fidelity” prototypes, without code, test the prototypes with users (users experience study) and use the findings to improve on the next iteration of the design leading to the design of a high-fidelity prototype. With these learning objectives, GenAI could be useful in brainstorming the problem identification, ideation of design scenarios for low-fidelity prototype, and generation of codes to model the high-fidelity prototype.

While our second case study of a graduate course in Data Visualization and Presentation is similar in that it focuses on the viewers’ perception and experience, it differs in that the writing of code is a primary learning objective. In this course, the process of producing graphics with ggplot2 (Wickham 2016), a graphical package within the tidyverse of R, is intrinsically linked with the pedagogical strategies employed. The syntax of ggplot2 code is fundamentally built on a layered grammar of graphics (Wickham 2010). This course emphasizes the metacognition of data visualization to help students build the habits of mind essential for the production of “graphical excellence” (Tufte 1983). Students become proficient in this by gaining exposure to a diverse array of published graphics and identifying graphical components. This cognitive mapping between physical aesthetics and the abstract organization of data, helps students to better understand how to collect, organize, and wrangle data so that it may be used for other downstream analyses. Although GenAI certainly can produce code in ggplot2, it cannot critique a graphics effectiveness for diverse audiences. However, there are opportunities for appropriate applications of GenAI, if prompted correctly. For instance, since GenAI relies on a vast body of scholarship, it can be prompted to effectively and efficiently create color palettes with considerations made for those with vision impairments and with cultural sensitivity for color connotation. This assistance from GenAI does not take away from the learning process, but instead provides a tool to increase accessibility.

## 6. Conclusion and Future Work

This study discussed the possibilities of developing equitable assessment strategies for STEM courses integrating the use of GenAI tools. Using a rapid review method, we provided an overview of literature designing assessment strategies with GenAI tools or for courses deploying GenAI as learning resources. Then we analyzed data generated through STEM subject experts' discussion on topics regarding assessment in the GenAI era. This study is limited in several ways. First, the rapid literature review, which is useful for gaining a quick overview of a field or area of study, could not provide an in-depth understanding of a phenomenon. Therefore, its findings cannot be generalizable.

Furthermore, subject experts' perspectives can be subjective and lack extensive coverage due to the limited number of experts who participate in this study. In the future, recruiting more experts for this type of study could be beneficial. While we did not claim an exhaustive approach nor generalization of this study's outcomes, its findings contribute to efforts towards advancing teaching and learning in a new era of GenAI. For example, the finding suggests that no consensus learning assessment strategy in a course using GenAI in a formal setting exists yet. We expect more scholarly discussions around this important topic. At the same time, this study contributes to the literature by suggesting various ways STEM educators could innovatively contextualize traditional strategies such as self-reflection and peer-reviewing to equitably assess students' learning outcomes in the era of GenAI. One of the areas in which contextualization of assessment strategies can be further developed is conducting a college-wide multidisciplinary study of faculty and students to analyze their perception of the use of GenAI tools in order to foster transparency and trust between teachers and students regarding the equitable assessment of learning outcomes.

## Acknowledgments

The authors gratefully acknowledge the support of a Title III SIP Grant, which funded the Summer Learning Circle. The Title III SIP Grant is funded by the United States Department of Education.

## References

Agbo, F. J. (2024, March). Broadening Participation in Adult Education: A Literature Review of Computer Science Education. In *Proceedings of the 55th ACM Technical Symposium on Computer Science Education V. 1* (pp. 11-17).

Algahtani, A. (2024). A Comparative Study of AI-Based Educational Tools: Evaluating User Interface Experience and Educational Impact. *Journal of Theoretical and Applied Information Technology, 102*(5).

Braun, V., & Clarke, V. (2023). Is thematic analysis used well in health psychology? A critical review of published research, with recommendations for quality practice and reporting. *Health Psychology Review*, *17*(4), 695-718.

Chau, R. C. W., Thu, K. M., Yu, O. Y., Hsung, R. T. C., Lo, E. C. M., & Lam, W. Y. H. (2024). Performance of generative artificial intelligence in dental licensing examinations. International Dental Journal, 74(3), 616-621.

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.

Dai, Y., Liu, A., & Lim, C. P. (2023). Reconceptualizing ChatGPT and generative AI as a student-driven innovation in higher education. Procedia CIRP, 119, 84-90

Drori, I., Zhang, S. J., Shuttleworth, R., Zhang, S., Tyser, K., Chin, Z., ... & Udell, M. (2023, August). From Human Days to Machine Seconds: Automatically Answering and Generating Machine Learning Final Exams. In *Proceedings of the 29th ACM SIGKDD Conference on Knowledge Discovery and Data Mining* (pp. 3947-3955).

Duane, A. (2024). Can Artificial Intelligence Complete My Assessment? A Student Led Initiative to Stress Test the Academic Integrity of University Assessment Using Generative AI. In the 10th International Conference on Higher Education Advances (HEAd'24). Editorial Universitat Politècnica de València.

Eva, T. P., Akter, S., Zabeen, M., & Shahriar, S. H. B. (2024). Exploring the future of learning: understanding the innovation in learning from the perspectives of developing nation. *Journal of Research in Innovative Teaching & Learning.*

Feng, T. H., Denny, P., Wuensche, B., Luxton-Reilly, A., & Hooper, S. (2024). More Than Meets the AI: Evaluating the performance of GPT-4 on Computer Graphics assessment questions. In Proceedings of the 26th Australasian Computing Education Conference (pp. 182-191).

Finnie-Ansley, J., Denny, P., Becker, B. A., Luxton-Reilly, A., & Prather, J. (2022, February). The robots are coming: Exploring the implications of openai codex on introductory programming. In Proceedings of the 24th Australasian Computing Education Conference (pp. 10-19)

Keane, B. The Learning Circle in culture change: why use it? Retrieved from https://www.pioneernetwork.net/wp-content/uploads/2016/10/The-Learning-Circle-in-Culture-Change.pdf

Khangura, S., Konnyu, K., Cushman, R., Grimshaw, J., & Moher, D. (2012). Evidence summaries: the evolution of a rapid review approach. Systematic reviews, 1, 1-9.
Khangura, S., Konnyu, K., Cushman, R., Grimshaw, J., & Moher, D. (2012). Evidence summaries: the evolution of a rapid review approach. Systematic reviews, 1, 1-9.
Koltovskaia, S., Rahmati, P., & Saeli, H. (2024). Graduate students' use of ChatGPT for academic text revision: Behavioral, cognitive, and affective engagement. *Journal of Second Language Writing, 65*, 101130.

Kohnke, L. (2024). Exploring EAP Students' Perceptions of GenAI and Traditional Grammar-Checking Tools for Language Learning. Computers and Education: Artificial Intelligence, 100279.

Law, L. (2024). Application of generative artificial intelligence (GenAI) in language teaching and learning: A scoping literature review. Computers and Education Open, 100174.

Lee, G. G., Latif, E., Wu, X., Liu, N., & Zhai, X. (2024). Applying large language models and chain-of-thought for automatic scoring. *Computers and Education: Artificial Intelligence, 6*, 100213.

Luo, J. (2024). How does GenAI affect trust in teacher-student relationships? Insights from students' assessment experiences. *Teaching in Higher Education,* 1-16.

Martínez-Téllez, R., & Camacho-Zuñiga, C. (2023, October). Enhancing mathematics education through AI chatbots in a flipped learning environment. In *2023 World Engineering Education Forum-Global Engineering Deans Council (WEEF-GEDC)* (pp. 1-8). IEEE.

Moorhouse, B. L., Yeo, M. A., & Wan, Y. (2023). Generative AI tools and assessment: Guidelines of the world's top-ranking universities. Computers and Education Open, 5, 100151.

Neve, A., Pawelczak, P. (2024) Using ChatGPT for Creating Computer Science Exercises – a Case Study from the Real World. 1*0th International Conference on Higher Education Advances (HEAd'24*). DOI: https://doi.org/10.4995/HEAd24.2024.17366

Ogunleye, B., Zakariyyah, K. I., Ajao, O., Olayinka, O., & Sharma, H. (2024). Higher education assessment practice in the era of generative AI tools. arXiv preprint arXiv:2404.01036.

Pinargote, A., Calderón, E., Cevallos, K., Carrillo, G., Chiluiza, K., & Echeverría, V. (2024). Automating Data Narratives in Learning Analytics Dashboards using GenAI. In LAK Workshops (pp. 150-161).

Radovi´c, S., Seidel, N., Haake, J. M., & Kasakowskij, R. (2023). Analyzing students' self-assessment practice in a distance education environment: Student behavior, accuracy, and task-related characteristics. Journal of Computer Assisted Learning. https://doi.org/10.1111/jcal.12907

Radovi´c, S., Firssova, O., Hummel, H. G. K., & Vermeulen, M. (2021). Improving academic performance: Strengthening the relation between theory and practice through prompted reflection. Active Learning in Higher Education, 1–15. https://doi.org/10.1177/14697874211014411

Roman, David. (2023). The history of Generative AI (GenAI). AI & ML Insights. Published on August 28, 2023. Accessed on August 16, 2024. https://wearebrain.com/blog/the-history-of-generative-ai-genai/

Sanusi, I. T., Oyelere, S. S., Agbo, F. J., & Suhonen, J. (2021, October). Survey of resources for introducing machine learning in K-12 context. In *2021 IEEE frontiers in education conference (FIE)* (pp. 1-9). IEEE.

Sanusi, I. T., Agbo, F. J., Dada, O. A., Yunusa, A. A., Aruleba, K. D., Obaido, G., … Centre for Multidisciplinary Research and Innovation (CEMRI). (2024). Stakeholders' insights on artificial intelligence education: Perspectives of teachers, students, and policymakers. Computers and Education Open, 7, 100212. doi:10.1016/j.caeo.2024.100212

Tan, C. W., & Lim, K. Y. (2023, October). Revolutionizing Formative Assessment in STEM Fields: Leveraging AI and NLP Techniques. In *2023 Asia Pacific Signal and Information Processing Association Annual Summit and Conference (APSIPA ASC)* (pp. 1357-1364). IEEE.

Tufte, E. (1983) The Visual Display of Quantitative Information. Graphics Press, Cheshire.

Wickham, H. (2010). A layered grammar of graphics. *Journal of computational and graphical statistics*, *19*(1), 3-28.

Wickham, H. (2016). ggplot2: Elegant graphics for data analysis. useR. Springer.

Wolfer, J. (2023, September). A Qualitative Assessment of ChatGPT Generated Code in the Computer Science Curriculum. In *the In*ternational *Conference on Interactive Collaborative Learning* (pp. 43-53). Cham: Springer Nature Switzerland.

Xia, Q., Weng, X., Ouyang, F., Lin, T. J., & Chiu, T. K. (2024). A scoping review on how generative artificial intelligence transforms assessment in higher education. International Journal of Educational Technology in Higher Education, 21(1), 40.

Zheng, C., Zhou, H., Meng, F., Zhou, J., & Huang, M. (2023, September). Large language models are not robust multiple choice selectors. In *The Twelfth International Conference on Learning Representations*.