2026 ASEE ANNUAL CONFERENCE June 21–24, 2026 | Charlotte Convention Center, NC
Host Sponsor UNC Charlotte
Paper ID #53410

The Impact of Guided GenAI Use and Paper-Based Testing on Student Engagement in HBCU Programming Course

Dr. Sm Zobaed, University of Maryland Eastern Shore
Dr. Weiwei Zhu Stone, University of Maryland Eastern Shore

Biographical Sketch Department of Computer Science and Engineering Technology University of Maryland Eastern Shore Weiwei Zhu, PhD

Devika Paranjpe, University of Maryland Eastern Shore

Abstract

This is an empirical full paper. The recent rapid growth of Generative AI (GenAI) imposes challenges to academic integrity and student engagement, especially in computer programming courses. This issue is alarming for most universities, particularly those at Historically Black Colleges and Universities (HBCUs). HBCU students often face unique psychosocial pressures that can lead to the abuse of AI to complete assignments which hinders the development of foundational computational thinking. To address these challenges, this study introduces a teaching methodology designed within an introductory programming course. The methodology consists of a two-pronged strategy: 1) transitioning online assessments to proctored paper-based tests to ensure academic integrity and 2) mandating the use of GenAI as a learning tool to engage students for a deeper understanding of class lectures and being well-prepared for tests. Our results indicate 75% increase in active student engagement in in-class participation, reflecting a significant positive shift in learning programming. We observed a 20% performance decline when transitioning from online to proctored assessments, revealing substantial academic integrity concerns with unsupervised formats. However, students demonstrated consistent recovery across subsequent proctored tests, achieving 22% improvement with a linear trend of $+3.6$ points per assessment. Based on our findings, this method offers a practical framework for educators to maintain academic rigor while exploiting the pedagogical benefits of AI which is particularly effective in supporting at-risk student populations.

1 Introduction

1.1 The Technological Disruption

This empirical, full paper investigates a pedagogical framework that balances the risks of GenAI-enabled academic dishonesty with its potential as a Socratic tutoring tool within an HBCU computer science context [1]. GenAI refers to a class of machine learning models, particularly Large Language Models (LLMs) that are engineered to synthesize high-quality text, imagery, and software code in response to natural language prompts. Since late 2022, the rapid integration of GenAI tools (e.g., ChatGPT by OpenAI, Claude by Anthropic, Gemini by Google *etc.*) into higher education has been praised for making expert-level teaching, learning, and research support accessible to all students [2, 3]. As an example, ChatGPT supports students in several ways: assisting with writing assignments, helping to clarify complex topics, and on demand real-time tutoring. Beyond this, it can generate personalized learning content and feedback, thereby increasing educational accessibility and adaptability [4, 5]. A recent research survey claimed that within months of public release of ChatGPT, nearly 70% of the surveyed students reported using AI tools for their coursework [6].

Any GenAI tools can offer immediate coding solutions that require no understanding of the underlying concepts. This is alarming for an introductory programming course (a.k.a. CS1) student struggling with the logic of conditional statements, loops, and defining functions. Educators across institutions have raised concerns about the authenticity of student submissions, and the evidence supports their worry. For instance, Finnie-Ansley et al. demonstrated that OpenAI's Codex performs at the level of a median CS1 student on examination questions [7]. Hence, the traditional paradigms of assessment and instruction in programming education are no longer tenable in their existing forms.

1.2 Crisis versus Opportunity

This disruption presents educators with a paradox. On the one hand, the ability of Gen AI to create complete solutions to programming assignments represents a threat to academic integrity. Students can bypass the productive struggle to reach solutions, which is essential in developing computational thinking. Every take-home assignment becomes a potential exercise in prompt engineering rather than programming. On the other hand, the same tools offer significant potential as teaching assistants. They explain error messages, trace through code execution, and answer follow-up questions with patience no human instructor can match at scale.

At HBCUs, this paradox takes on a particular sense of urgency. HBCUs have enabled African American students to advance in their education. Yet, many of the students that attend these colleges face multiple challenges. Financial constraints prevent many students from dedicating time to their education. The lack of preparation in mathematical and computational thinking further prevents students from completing their college education sucessfully [8]. Finally, many HBCU students carry the additional challenge of dealing with imposter syndrome and other similar psychological and sociological challenges [9]. In these conditions, the use of AI to generate passing work does more to undermine the development of genuine competency.

1.3 The Stakes: Equity and Representation

The stakes of this situation extend beyond individual grades. Computer science is a field with significant representation gaps – HBCUs have historically been crucial pipelines for diversifying the technology workforce. When students bypass foundational learning through AI dependency, they may achieve superficial success in introductory courses only to face insurmountable barriers in advanced coursework and professional practice. The short-term relief of AI-assisted assignment completion trades away long-term career readiness. Having students within an already marginalized population graduate without genuine competencies is fundamentally stereotypical and counterproductive to the diversification of computing fields. Therefore, the challenge here is not merely pedagogical but it has significant implications for equity and representation within an already marginalized population. Graduating without genuine competencies is fundamentally stereotypical and counterproductive to the diversification of computing fields. Therefore, the challenge here is not merely pedagogical but it has significant implications for equity and representation within the technology sector.

1.4 Contribution

This study proposes and evaluates a teaching methodology designed to navigate these competing tensions. The framework employs a two-pronged strategy that acknowledges both the threats and opportunities presented by GenAI. The first prong involves transitioning summative assessments from online formats to proctored, paper-based examinations. This shift removes the technological advantages that enable AI-assisted cheating during high-stakes evaluations, ensuring that the grades reflect authentic individual competency. The second prong mandates the structured use of GenAI as a learning tool during instruction and formative practice. Rather than prohibiting AI, which would be both futile and counterproductive given its ubiquity, this approach teaches students to use these tools as Socratic tutors that explain concepts and support deeper engagement with course material.

We observed roughly a 75% increase in classroom engagement after introducing structured AI tutoring. Exam scores dropped by at least 20% after the switch to proctored paper tests. We interpret that drop not as a failure of instruction but as a correction. Previous online exam scores were inflated. The paper-based format exposed where students actually stood. Early trends suggest scores are recovering as the benefits of genuine learning take hold.

1.5 Paper Organization

The following sections develop this framework in detail. Section 2 includes a review of the related literature on academic integrity in the age of AI, the use of large language models in education, and the challenges facing computer science education at minority-serving institutions. The methodology used in this research, the implementation of both paper-based and AI-assisted learning in the course sections included in this study are detailed in Section 3. Section 4 presents the findings from this research. Finally, Section 5 includes a discussion of the implications of these findings and suggestions for future research.

2 Related Work

2.1 Challenges and Risks of Generative AI in Education

The primary concern with integrating generative AI into education is academic dishonesty. Students may use ChatGPT to complete assignments or cheat on quizzes, undermining the educational process and devaluing academic credentials [10]. Cottonet al. highlight the dual nature of ChatGPT in academia, noting both integrity risks and engagement potential, and call for proactive measures such as AI recognition technologies, student education on ethical AI use, and explicit institutional policies [11]. Detection of AI-generated content has proven difficult [12, 13]. Farrelly and Baker found that commercial detection tools produce high false-positive rates, incorrectly flagging human-written text as AI-generated, with particularly poor performance on non-native English writing [14]. This makes detection-based enforcement both unfair and unreliable. Kovari similarly reported that traditional plagiarism tools cannot identify text produced by advanced models like ChatGPT [10]. Institutional policies remain inconsistent, leaving students confused about what is permitted. Lau and Guo found that programming instructors' responses ranged from outright prohibition to full integration, often varying within the same department [15]. Institutions must therefore implement clear policies, design assessments requiring critical thinking, and provide training on ethical AI use.

2.2 AI as a Pedagogical Tool

While generative AI raises concerns about misuse, it also holds promise for improving learning outcomes. Kasneciet al. reviewed applications of large language models to education and found several productive ways to use such systems - to answer questions from students, to provide worked examples, and to offer feedback [16]. Unlike earlier intelligent tutoring systems that required extensive development for particular domains, current models work across subjects without customization [17]. MacNeilet al. incorporated AI-generated code explanations into a web development textbook. Students utilizing these explanations developed a better understanding of key concepts than those who did not utilize them [18]. The distinction here was that students employed the AI to understand existing code rather than generate new code. This same pattern emerges throughout the literature [19, 20]. Students who struggle with problems prior to receiving instruction retain more knowledge than those exposed to instruction first. Tools and features within AI systems that eliminate difficulties by providing instant answers do interfere with learning - tools and features that support difficulty in learning actually enhance it. Holmeset al. make clear that AI should enhance rather than replace crucial human elements of teaching such as mentorship and ethical judgment [21].

2.3 Current Strategies to Counter Unethical AI Use

The rise of LLMs in education has led many educators and institutions to develop approaches to prevent misuse while adapting to AI-enhanced learning environments. Several institutions have established clear policies on AI usage that require students to disclose when they have used AI tools in their assignments [22]. In addition, they have implemented AI-detecting tools within plagiarism-checking programs; services such as Turnitin [1] have introduced algorithms to detect AI-generated text. Educators are designing assessments that require high levels of originality and creativity for which AI tools are less effective, such as personal reflective assignments or those requiring specific local contexts.

Moreover, some educators have adopted oral exams in which students must present and defend their work, making it difficult to rely on AI-generated content. Fifth, collaborative group work and peer review mechanisms increase the chances of identifying inconsistencies or potential AI misuse. Empirical evidence supports the effectiveness of adaptive and reflective evaluation. Moorhouse et al. highlight successful pilot programs at highly regarded institutions that incorporate reflective and personalized tasks, limiting AI misuse by requiring individualized responses [23]. In [24], Dempere et al. depict that ethical AI use campaigns combined with AI recognition technologies greatly improve academic integrity compliance. However, upon evaluating AI tools, we observe that they are not capable of detecting AI-generated code. Subsequently, the integrity of the CS1 quiz and exam submissions still remains questionable. To this end, we realize the necessity of human-proctored paper-based exams.

[1] https://www.turnitin.com

2.4 CS1 at Minority-Serving Institutions

Introductory programming courses have high attrition rates across all institution types. A number of research reviewed decades of CS1 research and found failure and withdrawal rates consistently between 15% to 20% [25, 26]. The causes are multiple and convoluted which makes the problem resistant to simple interventions. At HBCUs and other minority-serving institutions, additional factors apply. For example, their students often have less prior programming experience that indicates unequal access to computer education in grades K-12 rather than differences in aptitude [27]. In [28], Charleston et al. described the psychological pressures upon Black students within computing programs - stereotype threat, imposter syndrome and the burden of representing communities. Such pressures tend towards discouragement of help-seeking behaviour and reliance upon shortcuts. Effective interventions must take into account both skills and confidence. Approaches to instruction that build genuine competence achieve both cognitive and affective goals. A student who masters foundational concepts has both the knowledge necessary for success in more advanced courses and the confidence to believe in their belonging within the field. Hence, institutions ensure that AI complements traditional learning in upholding academic integrity and critical thinking.

3 Methodology

3.1 Course and Participants

We implemented this intervention in an introductory-level programming course entitled "Python Programming" within an HBCU during the Fall 2025 semester. The course covers the standard content associated with an introductory level course - variables, conditionals, loops, functions, lists, and GUI basics - all within Python. Students attend a 50-minute lecture four days per week. Prior experience in programming within the cohort ranges from none to moderate. Prior programming experience across the cohort ranges from none to moderate.

3.2 Strategy One: Paper-Based Proctored Exams

Previously, exams were administered through Canvas, the university's LMS. Students completed tests on personal devices, sometimes in computer labs or remotely. Although academic integrity policies prohibit unauthorized assistance, enforcement was effectively impossible under these conditions. A student could simply open AI tools in a second browser tab without detection. We moved all major exams to a paper-based, proctored format. Students receive printed booklets and write code by hand. Electronic devices are collected before testing begins. Instructors and teaching assistants monitor the room throughout. This aligns with Kovari's recommendation that real-time proctored exams reduce the likelihood of AI-assisted assessment completion [10].

We recognize that paper-based exams remove two advantages simultaneously: AI assistance and real-time compiler feedback. A student writing code on a computer, even without AI, can run their program, catch typos, and iterate on error messages. Handwritten code offers none of that. To account for this, our grading scheme distinguishes between conceptual errors and surface-level syntax mistakes. Minor typos, missing colons, or small indentation issues do not cost points as long as the underlying logic is correct and the student's intent is clear. This policy is communicated before each exam so students can focus on problem solving rather than exact syntax. We chose this approach over alternatives such as restricted-internet computer exams because browser-level restrictions are difficult to enforce reliably across student devices, and AI tools can still be accessed through phones or secondary devices. Collecting all electronics and using printed materials was the most practical way to ensure a controlled environment. The loss of compiler feedback is a known cost, but our grading adjustments ensure students are not penalized for the absence of a feature unrelated to their conceptual understanding.

3.3 Strategy Two: Structured AI Tutoring

Banning AI outright would be unenforceable and counterproductive. The tools are freely available to all students, and they will show up in the students' development environments regardless of the course policy. The question is not whether the students use AI but how they use it. We have structured our courses to ask the AI tools to explain code rather than write it.

During the first week of the semester, we taught the students how to use AI tools like ChatGPT as tutors for their programming challenges. The students can pose questions regarding the code with which they are struggling. However, pasting homework assignments and submitting entries generated by such AI tools amounts to academic dishonesty and will be dealt with in accordance with the school code of conduct.

This strategy is built into the course workflow. Before each lecture, the students are given an assignment to review the lecture topic with the AI tool of their choosing and submit their reflections on the topic. Some in-class exercises ask the students to pose questions regarding topics they are struggling with, use the AI tool to find the answer and present that to the rest of the class. Using these strategies the students begin to learn how to ask for help with their challenges and assess the answers generated by these AI tools - while recognizing that the tools are not infallible.

3.4 Theoretical Framework and Hypotheses

The proposed methodology draws upon principles from recent literature regarding AI integration within education. Such principles include considerations of academic integrity [10, 11], critical thinking and human-AI collaboration [21, 16], equitable access in alignment with educational goals [28, 27], data privacy [22] and AI literacy [23, 24]. Each of these aspects can then be operationalized within the specific context of an introductory programming course within such an HBCU.

We expect three outcomes:

- Classroom engagement should increase. Students who practice asking questions to a non-judgmental AI may become more willing to participate in class discussion. This hypothesis draws on research showing that AI interaction removes the social cost of confusion, particularly for students who experience stereotype threat.
- Initial exam scores should decline. Earlier grades likely reflected AI assistance, and removing that assistance would expose the actual baseline of student understanding.

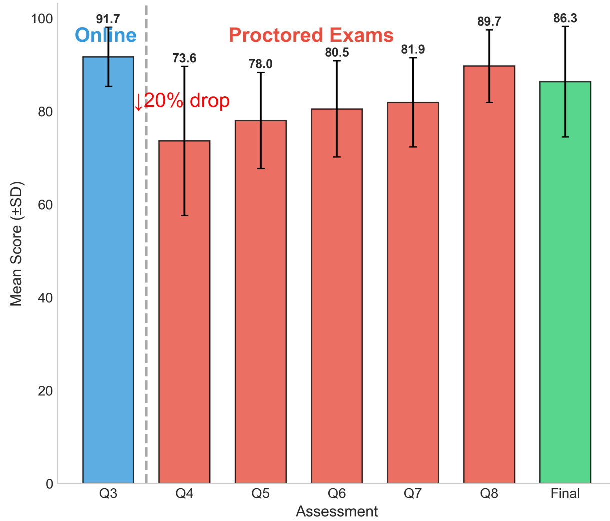


Figure 1: Performance trajectory showing 20% decline when transitioning from online (Q3, blue) to proctored assessments (Q4-Q8, red), followed by 22% recovery. Error bars represent standard deviation.

- Exam scores should recover over subsequent assessments as genuine learning accumulates through structured AI tutoring and as students recognize that dishonesty is no longer a viable strategy. This recovery, if observed, would itself serve as evidence that the methodology builds real competence over time.

4 Experimental Results

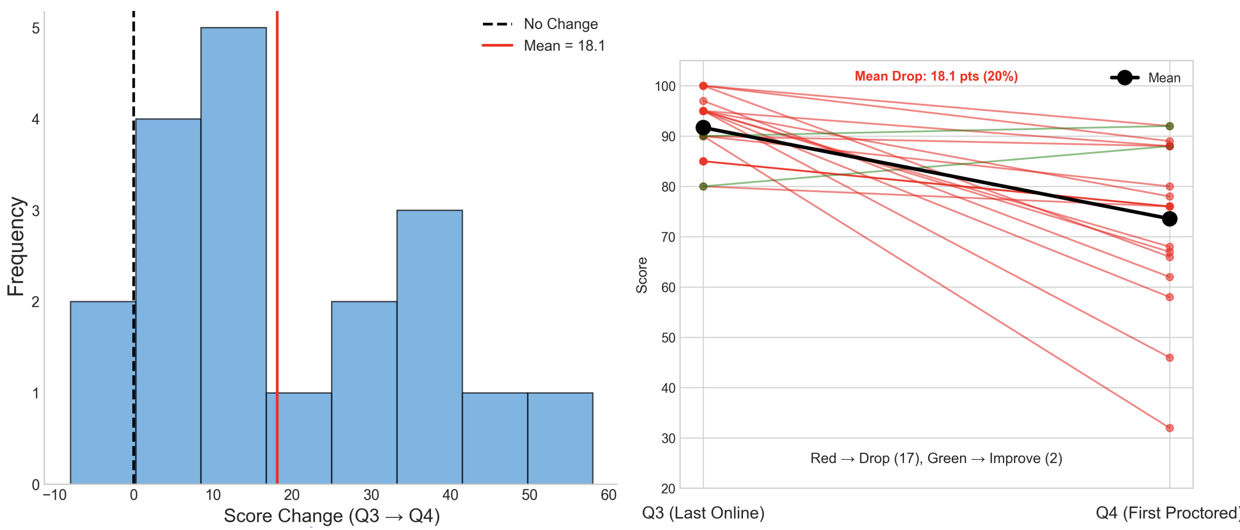
This section presents findings from our two-pronged pedagogical intervention combining mandatory GenAI usage with proctored paper-based assessments. Data were collected from 19 students across nine assessments during Fall 2025: three online unproctored quizzes (Q1-Q3, Weeks 1-7), five in-person proctored quizzes (Q4-Q8, Weeks 8-16), and one comprehensive proctored final exam (Week 17). Note that unproctored quizzes were still proctored by software called *Yuja* but students realized how to dodge the software. Therefore, in our context, "proctor" refers specifically to human supervision. Statistical analyses employed paired t-tests and Cohen's d effect sizes with significance threshold $\alpha = 0.05$.

4.1 Immediate Impact of Transitioning to Proctored Assessments

The transition from online to proctored assessments revealed substantial performance decline, providing evidence of academic integrity concerns. Figure 1 presents the complete performance trajectory across all nine assessments. The mean score for Q3, the final online assessment, was M = 91.7 (SD = 6.8). Upon implementing the first proctored assessment (Q4), mean performance dropped to $M = 73.6$ $(SD = 11.5)$, representing a 20% decline from Q3. This difference was highly significant ($t(18) = 5.89, p < 0.001$, $Cohen's\ d = 1.95$).

Figure 2 illustrates the distribution and individual trajectories of score changes from Q3 (last online exam) to Q4 (first proctored exam). The histogram in Figure 2a shows the distribution of score changes across all students, with a mean drop of 18.1 points indicated by the red vertical line. The distribution is positively skewed, with the modal category falling in the 10–20 point decline range (frequency = 5). Although most students cluster in the drop range of 0-20 points, a notable tail extends toward larger declines of 40–60 points, suggesting that some students experienced particularly severe performance drops under proctored conditions

Individual-level analysis reported in Figure 2b, reveals that over 87% students experienced performance declines ranging from 5 to 68 points (median = 17 points), while only 2 students (10.5%) showed improvements. The magnitude of decline is independent of Q3 performance ($r = -0.18, p = 0.47$), indicating systemic reliance on unauthorized resources rather than selective misconduct. This decline underscores that online assessment scores substantially overestimated actual programming competencies.



(a) Individual Score Changes from Q3 (online) → Q4 (first proctored exam).

(b) Individual-level transition analysis from Q3 → Q4.

Figure 2: Comparison of individual-level transitions and score changes from Q3 to Q4.

4.2 Recovery Patterns and Skill Development

According to Figure 3, after the initial decline, the students demonstrated a consistent improvement in performance on the proctored assessments. Mean scores increased progressively: Q4 ($M = 73.6, SD = 16.0$), Q5 ($M = 78.0, SD = 10.3$), Q6 ($M = 80.5, SD = 10.3$), Q7 ($M = 81.9, SD = 9.6$), and Q8 ($M = 89.7, SD = 7.8$).

From Q4 to Q8, students achieved a total recovery of 16.1 points (22% improvement). Linear regression revealed a significant positive trend with $slope = +3.6$ points per assessment ($R^2 = 0.93, p = 0.008$), as shown by the blue dashed trend line in the figure. The Q4 to Q8 improvement was statistically significant ($t(18) = -4.30, p < 0.001, Cohen's\ d = 0.99$), representing a large effect size. Additionally, SD decreased from 16.0 to 7.8 points, indicating increasing consistency as competencies solidified.

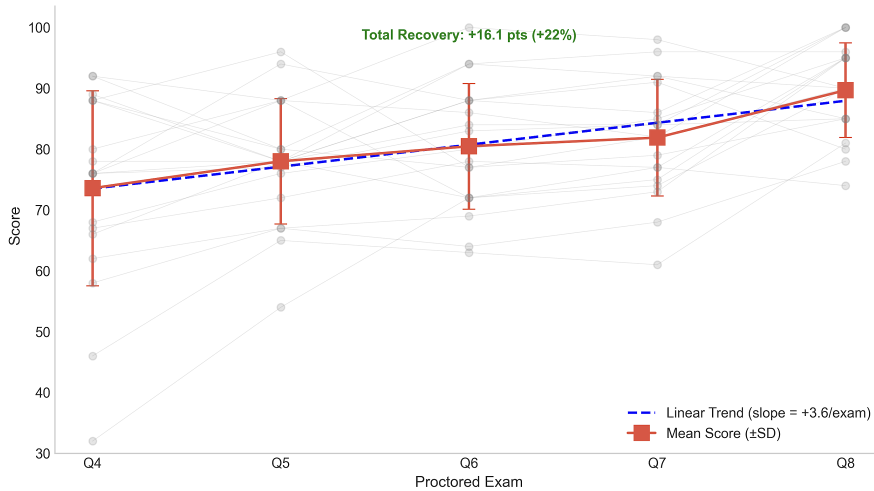


Figure 3: Recovery pattern across five proctored exams (Q4-Q8). Gray lines show individual trajectories and red squares represent mean scores with SD.

A critical question is whether this recovery reflects genuine skill acquisition or merely increased familiarity with the proctored testing format. To address this concern, we examined the predictive validity of assessment scores by correlating them with final exam performance. If students developed genuine programming competencies, their proctored exam scores should predict final exam success. As illustrated in Figure 1, the performance of the final exam $(M = 86.3, SD = 11.9)$ is strongly correlated with the mean of the test administered $(r = 0.74, p < 0.001)$ which explains 54% of the variance. In contrast, the online assessment (Q3) did not show significant correlation with the final exam $(r = -0.19, p = 0.44)$. This divergent pattern provides evidence that proctored assessments captured genuine understanding that transferred to independent performance on the final exam.

4.3 Student Engagement Metrics

Following implementation of mandatory GenAI usage (Week 8), we observed substantial increases in engagement compared to baseline (Weeks 1-7) which are as follows:

- **In-class questions:** Increased 75%
- **Voluntary code demonstrations:** Increased 83%
- **Pre-quiz Assignment:** 100% completion rate. Students reported that mandatory GenAI engagement for homework enabled them to clarify conceptual confusion before class, facilitating more sophisticated in-class discussions.
- **Class attendance:** Increased from 78% to 96% (ignored dropped students)
- **Critical thinking:** Improved substantially as students' questions were evolved from procedural clarifications to higher-order thinking (e.g., ”Why does this solution work?”, ”How can I optimize it?”)

5 Conclusion

This study investigates the impact of a two-pronged pedagogical intervention - guided GenAI use combined with paper-based testing in an introductory programming course at an HBCU. Our findings demonstrate that this approach substantially enhances student engagement while maintaining academic integrity. The transition from online to proctored assessments revealed a 20% performance decline affecting 90% of students that provides empirical evidence of systemic academic integrity concerns with unsupervised online formats.

The incorporation of GenAI led to a 75% increase in the number of questions asked in class, an 83% increase in the number of students who voluntarily demonstrated what they learned in the lessons, and a rise in attendance from 78% to 96%. Additionally, there was a shift in the type of questions asked of students by lower-middle performers, indicating that they understood the concepts rather than merely memorizing processes. This shift signals that the GenAI tools helped these students to learn the material.

5.1 Implications and Recommendations

Our findings offer practical guidance for educators at HBCUs and similar institutions:

- **Academic Integrity:** Implementing proctored tests will ensure that the measurement of student learning is accurate. Administering unsupervised and online tests will distort the measurement of a student's knowledge.
- **GenAI as a Pedagogical Tool:** Requiring the use of GenAI for learning will force students to engage with the technology. Knowing that proctored tests will be used to test their skills will encourage students to use these AI tools effectively.
- **Supporting At-Risk Students:** Seeing that at-risk students benefited the most suggests that this approach can effectively be used to help at-risk students succeed.

5.2 Limitations and Future Work

The limitations of this study include the small size of the test group ($N = 19$), the timeframe in which the research was conducted (one semester), and the number of institutions involved in the research. Future studies could focus on conducting longitudinal studies, testing different strategies for using GenAI in the classroom, obtaining qualitative data about student perceptions of engagement, and working to make the model scalable to multiple HBCUs.

In the age of GenAI, people face both challenges and opportunities in integrating AI into their fields of knowledge and expertise. The findings of this study indicate that incorporating guided use of GenAI alongside paper tests enables students to become more actively involved in the study. This outcome could enable HBCU faculty to increase student engagement in the classroom and better prepare students for future professional environments.

References

[1] J. Tang, L. Xia, Z. Li, and C. Huang, "Ai-researcher: Autonomous scientific innovation," *arXiv preprint arXiv:2505.18705*, 2025.

[2] P. S. Bhullar, M. Joshi, and R. Chugh, "Chatgpt in higher education-a synthesis of the literature and a future research agenda," *Education and Information Technologies*, vol. 29, no. 16, pp. 21 501–21 522, 2024.

[3] D. Lee, M. Arnold, A. Srivastava, K. Plastow, P. Strelan, F. Ploeckl, D. Lekkas, and E. Palmer, "The impact of generative ai on higher education learning and teaching: A study of educators' perspectives," *Computers and Education: Artificial Intelligence*, vol. 6, p. 100221, 2024.

[4] I. Celik, "Towards intelligent-tpack: An empirical study on teachers' professional knowledge to ethically integrate artificial intelligence (ai)-based tools into education," *Computers in human behavior*, vol. 138, p. 107468, 2023.

[5] M. Montenegro-Rueda, J. Fernández-Cerero, J. M. Fernández-Batanero, and E. López-Meneses, "Impact of the implementation of chatgpt in education: A systematic review," *Computers*, vol. 12, no. 8, p. 153, 2023.

[6] D. Smerdon, "Ai in essay-based assessment: Student adoption, usage, and performance," *Computers and Education: Artificial Intelligence*, vol. 7, p. 100288, 2024.

[7] J. Finnie-Ansley, P. Denny, A. Luxton-Reilly, E. A. Santos, J. Prather, and B. A. Becker, "My ai wants to know if this will be on the exam: Testing openai's codex on cs2 programming exercises," in *Proceedings of the 25th Australasian Computing Education Conference*, 2023, pp. 97–104.

[8] K. Mills, M. Coenraad, P. Ruiz, Q. Burke, and J. Weisgrau, "Computational thinking for an inclusive world: A resource for educators to learn and lead." *Digital Promise*, 2021.

[9] L. Charleston, R. P. Adserias, N. M. Lang, and J. Jackson, "Intersectionality and stem: The role of race and gender in the academic pursuits of african american women in stem," *Journal of Progressive Policy & Practice*, vol. 2, no. 3, pp. 273–293, 2014.

[10] A. Kovari, "Ethical use of chatgpt in education—best practices to combat ai-induced plagiarism," in *Frontiers in Education*, vol. 9. Frontiers Media SA, 2025, p. 1465703.

[11] D. R. Cotton, P. A. Cotton, and J. R. Shipway, "Chatting and cheating: Ensuring academic integrity in the era of chatgpt," *Innovations in education and teaching international*, vol. 61, no. 2, pp. 228–239, 2024.

[12] R. Kumar and M. Mindzak, "Who wrote this? detecting artificial intelligence–generated text from human-written text," *Canadian Perspectives on Academic Integrity*, vol. 7, no. 1, 2024.

[13] S. Abdali, R. Anarfi, C. Barberan, and J. He, "Decoding the ai pen: Techniques and challenges in detecting ai-generated text," in *Proceedings of the 30th ACM SIGKDD Conference on Knowledge Discovery and Data Mining*, 2024, pp. 6428–6436.

[14] T. Farrelly and N. Baker, "Generative artificial intelligence: Implications and considerations for higher education practice," *Education Sciences*, vol. 13, no. 11, p. 1109, 2023.

[15] S. Lau and P. Guo, "From" ban it till we understand it" to" resistance is futile": How university programming instructors plan to adapt as more students use ai code generation and explanation tools such as chatgpt and github copilot," in *Proceedings of the 2023 ACM Conference on International Computing Education Research-Volume 1*, 2023, pp. 106–121.

[16] E. Kasneci, K. Seßler, S. Küchemann, M. Bannert, D. Dementieva, F. Fischer, U. Gasser, G. Groh, S. Günnemann, E. Hüllermeier *et al.*, "Chatgpt for good? on opportunities and challenges of large language models for education," *Learning and individual differences*, vol. 103, p. 102274, 2023.

[17] A. M. Bettayeb, M. Abu Talib, A. Z. Sobhe Altayasinah, and F. Dakalbab, "Exploring the impact of chatgpt: conversational ai in education," in *Frontiers in Education*, vol. 9. Frontiers Media SA, 2024, p. 1379796.

[18] S. MacNeil, A. Tran, A. Hellas, J. Kim, S. Sarsa, P. Denny, S. Bernstein, and J. Leinonen, "Experiences from using code explanations generated by large language models in a web software development e-book," in *Proceedings of the 54th ACM Technical Symposium on Computer Science Education V. 1*, 2023, pp. 931–937.

[19] A. D. Damiano, E. J. Lauría, C. Sarmiento, and N. Zhao, "Early perceptions of teaching and learning using generative ai in higher education," *Journal of Educational Technology Systems*, vol. 52, no. 3, pp. 346–375, 2024.

[20] I. J. Akpan, Y. M. Kobara, J. Owolabi, A. A. Akpan, and O. F. Offodile, "Conversational and generative artificial intelligence and human–chatbot interaction in education and research," *International Transactions in Operational Research*, vol. 32, no. 3, pp. 1251–1281, 2025.

[21] W. Holmes, M. Bialik, and C. Fadel, *Artificial intelligence in education promises and implications for teaching and learning*. Center for Curriculum Redesign, 2019.

[22] C. K. Y. Chan, "A comprehensive ai policy education framework for university teaching and learning," *International journal of educational technology in higher education*, vol. 20, no. 1, p. 38, 2023.

[23] B. L. Moorhouse, Y. Wan, C. Wu, L. Kohnke, T. Y. Ho, and T. Kwong, "Developing language teachers' professional generative ai competence: An intervention study in an initial language teacher education course," *System*, vol. 125, p. 103399, 2024.

[24] J. Dempere, K. Modugu, A. Hesham, and L. K. Ramasamy, "The impact of chatgpt on higher education," in *Frontiers in Education*, vol. 8. Frontiers Media SA, 2023, p. 1206936.

[25] J. Bennedsen and M. E. Caspersen, "Failure rates in introductory programming: 12 years later," *ACM inroads*, vol. 10, no. 2, pp. 30–36, 2019.

[26] C. Watson and F. W. Li, "Failure rates in introductory programming revisited," in *Proceedings of the 2014 conference on Innovation & technology in computer science education*, 2014, pp. 39–44.

[27] S. Mackay and A. Decker, "Updating our understanding of the impact of pre-college computing experiences on university students," in *2020 IEEE Frontiers in Education Conference (FIE)*. IEEE, 2020, pp. 1–5.

[28] L. J. Charleston, P. L. George, J. F. Jackson, J. Berhanu, and M. H. Amechi, "Navigating underrepresented stem spaces: Experiences of black women in us computing science higher education programs who actualize success." *Journal of Diversity in Higher Education*, vol. 7, no. 3, p. 166, 2014.