An Exploratory Study on the Impact of AI tools on the Student Experience in Programming Courses: an Intersectional Analysis Approach

Mary Lou Maher
University of North Carolina, Charlotte
m.maher@charlotte.edu

Sri Yash Tadimalla
University of North Carolina, Charlotte
stadimal@charlotte.edu

Dhruv Dhamani
University of North Carolina, Charlotte
ddhamani@charlotte.edu

2023 IEEE Frontiers in Education Conference (FIE) | 979-8-3503-3642-9/23/$31.00 ©2023 IEEE | DOI: 10.1109/FIE58773.2023.10343037

***Abstract*— This work-in-progress paper presents a study that sheds light on the concerns that students may not develop sufficient programming skills and as a result, be less competent with the use of ChatGPT. The potential benefits for students are significant: Access to ChatGPT increases the ability for students to work constructively on their own schedule. The ease of use of ChatGPT may engage students who might otherwise hesitate in asking for support. Before these tools can be meaningfully introduced into a course, work must be done to study the impact of these AI tools on a student's ability to learn. In this study, participants are recruited from introductory Java programming courses at a large public university in the United States. This paper presents preliminary findings from a mixed method study design that consists of a pre-task assessment quiz; and a programming task in one of three conditions: (1) with no external help, (2) with the help of an AI chatbot, or (3) with the help of a generative AI tool like GitHub Copilot; followed by a post-task assessment and an interview on their experience and perceptions of the tools. Our preliminary findings describe our data collection, thematic analysis of the students' prompts and chatGPT responses, and a summary of the experience for 3 students. Our findings demonstrate a range of students' attitudes and behaviors towards chatGPT that provides insight for future research and plans for incorporating such AI tools in a course.**

***Index terms*—ChatGPT, CS education, Student Experience, Intersectionality**

I. Introduction

The recent availability of Large Language Models (LLMs) like ChatGPT that respond to natural language prompts with answers that can include Java code raises concerns and questions on how its use will affect learning. Educational institutions are considering and developing policy and ethical issues regarding its use in assignments and exams. There are also recognized benefits to integrating AI tools into courses. Real-time, 24/7 access to ChatGPT support may foster self-paced learning and potentially broaden the success spectrum in a CS major by facilitating accessibility for a diverse student population [1]. However, before introducing ChatGPT-like tools into a course, it is important to study their impact on student learning. In this paper, we delve into the intricacies of how students engage with ChatGPT during programming tasks, presenting both empirical observations and thematic analysis of their interactions, as we explore the competency-building and cognitive experience of students using ChatGPT while performing a programming task.

While LLMs have been used in Natural Language Processing for decades, recent advances in deep learning and neural network architectures have led to the development of LLMs that exceed the capabilities of previous models [2]. These LLMs have demonstrated remarkable abilities, including general linguistic intelligence and unsupervised multitask learning [3]. LLMs are considered to be "general purpose" models and can even train themselves based on their use [4]. In this context, LLM assistance in programming can resemble a highly intelligent and flexible compiler or a partner in pair programming, presenting new challenges and opportunities for research in human-centric programming. Studies found that students preferred LLM-powered tools over other tools that provide code completion, [5].

While there are clear benefits to using AI tools in educational contexts, it is important to understand the potential impact on student learning. Several studies have previously examined the effectiveness of AI tools for enhancing student learning outcomes, such as increased engagement, improved knowledge retention, and enhanced problem-solving abilities [6, 7]. However, AI tools that can independently solve introductory CS programming exercises and are pervasively available have only emerged recently [8, 9, 10, 11]. There is a potential concern that heavy reliance on ChatGPT-like tools during formative assessments might compromise the development of essential programming skills and critical thinking abilities [11].

Therefore, it is crucial for researchers and educators to explore the impact of AI tools on student learning in programming courses through rigorous empirical studies. This will allow for a better understanding of how AI tools can be optimally integrated into programming courses to enhance student learning outcomes.

Previous studies of students using AI tools have mainly focused on the tool's performance, the test-taking ability of the models, or the interface usability. In contrast, our study addresses the following research question:

- How do students engage with ChatGPT in the context of programming tasks, and what impact does this have on their learning outcomes?

In this paper, we detail our study design in the 'Methodology' section, present preliminary findings in the 'Findings' section, reflect on our results in the 'Discussion' section, and conclude with implications and future directions.

II. Methodology

A mixed-method study design was developed to capture the attitudes and behaviors of students towards AI tools when working on a competency-building exercise similar to those found in an introductory Java programming course. The design includes a survey that captures student identity; a pre-task assessment quiz; and a programming task in one of three conditions: (1) with no external help, (2) with the help of an AI chatbot, or (3) with the help of a generative AI tool like GitHub Copilot; a post-task assessment quiz; and an interview on their experience and perceptions of the tools. Although the larger study utilizes a between-subjects design where each participant is subjected to only one of the three conditions, this work-in-progress paper focuses on preliminary findings from three diverse students who interacted with the ChatGPT tool.

Data Collection Instruments - The study includes three categories of data collection instruments to provide sufficient data to enable a more case study-like presentation of the findings: The first category is data collected about the student's identity and knowledge about the Java programming concepts being exercised in the programming tasks. The second category is data about their use of ChatGPT while performing the programming tasks. The third category is data about their knowledge of the same programming concepts assessed before the programming task and their reflection on the use of ChatGPT. A summary of the data collection instruments is:

Before the programming tasks:

- A survey that includes questions about the participant's attitude towards AI and demographic data that enables an intersectional analysis of each participant.
- Pre-task knowledge quiz in the form of multiple choice questions about the Java concepts of loops and arrays.

While performing the tasks with chatGPT:

- Observations of the participants' behavior while performing tasks and using ChatGPT.
- Transcripts of the prompts to chatGPT and responses from ChatGPT.
- A recording of the session that provides data such as the time spent on different activities, text interactions with ChatGPT, and a record of satisfactory or unsatisfactory completion of the programming tasks.

After performing the tasks:

- Post-task knowledge quiz with the same questions as the pre-task quiz.
- A recording and transcription of the interview that explores the participants' experience using the AI tools during the programming tasks.

Guiding questions include topics such as the helpfulness and role of ChatGPT, whether they learned anything about programming in Java from using ChatGPT, and if the AI tool provided support or challenge.

Data Analysis - Quantitative data was used to indicate comparisons in the amount of time each participant spent on the programming tasks and the number of prompt-response pairs and length of each prompt and response in characters. This quantitative data shows differences across the 3 participants in the ways they used ChatGPT.

A qualitative analysis was performed on the transcripts of the prompts and responses using a thematic analysis to systematically identify, organize, and offer insight into patterns of meaning (themes) across the data set [12]. A key data-organizing structure in the thematic analysis is coding. "A code in qualitative inquiry is most often a word or short phrase that symbolically assigns a summative, salient, essence-capturing, and/or evocative attribute for a portion of language-based or visual data" [13].

The Thematic Analysis of Student Interviews - The process involved the analyses of data (the scripts from the interviews) and coding, where the codes emerge from the data. The data analysis phase includes repeated comparisons between emerging codes and labels developed. Thus each participant's interview was coded via selective coding independently, applying qualitative verification strategies to compare codes, resulting in the development of thematic labels, followed by a derivation of overarching themes across the dataset.

The Thematic Analysis of ChatGPT interactions - The thematic analysis was performed by two independent coders, followed by a consensus. The three authors discussed the resulting codes and themes until agreement was

achieved. The overarching themes for the prompt-response data are the content of the prompt-response pair and the evaluation of the prompt-response pair. [14]. The themes for the content comprise: learning new knowledge and aligning the AI response. Learning new knowledge was coded as learning a concept, learning about Java implementation, or learning about debugging. Aligning the AI is described as setting the context or creating a persona for the AI. The themes for evaluation comprise the quality of the prompt-response pair and the type of iteration for the prompt-response pair. Codes for the quality of the prompt-response pair are reasonable prompt, reasonable response, and an indicator of the correctness of the response. Codes for the type of iteration are an indicator of the continuation from the previous prompt-response pair, a comparison with the response expected from a human tutor, and an indicator of whether the participant found the response helpful. The themes and codes provided the basis for the findings as a description of each participant's interaction, attitude, and behavior with ChatGPT.

III. Findings

Our findings are presented as a description of the experience of each participant.

Participant 1 was highly confident in Java programming, expressed an advanced understanding of Java's array and loop structures, and was capable of handling complex exercises related to these topics. Their previous experience with AI tools, such as ChatGPT, was evident during the study. Currently, they are pursuing a CS3 course and have a year's worth of experience in Java Programming. Participant 1 is an 18-20-year-old Asian student from a Vietnamese background. They identify as male and as a member of the LGBTQIA+ community. Their educational background does not specify whether they are a first-generation student or receiving financial aid. Participant 1 has worked at a job to support themselves and/or their family while in college.

The pre-task quiz was completed in 2 minutes, scoring 4 out of 5, with a minor error in a question on while loops.

Programming Task A, involving calculating array averages, took 4 minutes, completed without ChatGPT assistance. Task B, reversing a long value, required 10 minutes, with the participant expertly interacting with ChatGPT to answer Java programming questions. Task C, forming a Pascal triangle of a given size, took 17 minutes, with the participant employing ChatGPT for conceptual questions, algorithm design, troubleshooting, and understanding compiler errors and logic errors.

The post-task quiz was completed flawlessly in 2 minutes.

During the interview, Participant 1 praised ChatGPT's assistance in programming tasks and recognized its potential as a learning tool. However, they cautioned against overreliance on AI tools and the risk of hampering learning by simply copying templates. They expressed interest in AI alignment and safety, noting the importance of AI tools in enhancing programming comprehension, despite concerns about their impact on education and professional environments.

Participant 2 reported high confidence in Java programming and demonstrated a strong understanding of arrays and loops in Java. They are currently enrolled in a CS2 course and have been programming in Java for 6 months. However, their limited prior experience with AI tools was evident. They are an 18-20-year-old Black/African-American male from a small city or town. Participant 2 is a first-generation student receiving financial aid, including grants, loans, federal work-study, and scholarships. They work part-time to support themselves and/or their family, either on or off campus.

The pre-task quiz was completed in 4 minutes with a score of 4 out of 5, with a minor error in array creation syntax.

In Task A, they spent 5.5 minutes and consulted ChatGPT for array length determination. Task B required 11 minutes, with the participant using ChatGPT to ask multiple poorly constructed questions on algorithm design and Java programming. Despite an initially unclear query, rephrasing led to a more useful response. Task C, however, remained incomplete after 26 minutes, despite multiple queries to ChatGPT.

The post-task quiz took 3 minutes, with a score of 4 out of 5. They got the same question wrong again.

In the interview, Participant 2 acknowledged the slight improvement in their understanding of Java loops and arrays, attributing it to ChatGPT's assistance. They also voiced concerns about the potential misuse of AI tools in classroom settings but didn't see any issues with their professional use.

Participant 3 expressed average confidence in Java programming, and showed a good understanding of arrays and loops in Java, despite their low experience with AI tools. Currently, they are pursuing a CS3 course. Participant 3 is a white male aged between 21-24 from a small city in the US. They are not a first-generation student and are receiving loans as financial aid for their college education. They work part-time to support themselves financially.

The pre-task quiz was completed in 2 minutes 15 seconds, scoring 4 out of 5 due to a minor error in array creation syntax.

Task A was completed in 3 minutes 45 seconds independently. Task B took 7 minutes, with Participant 3 utilizing

ChatGPT effectively to understand long reversal. Task C took 8 minutes, with the participant using ChatGPT for problem-solving after the initial challenges. They made heavy use of ChatGPT, and used ChatGPT solutions for both Task B and C. Participant 3 made no attempt to further understand the solution they submitted, but when asked to review the solution during the interview, they were able to adequately explain both solutions.

The post-task quiz was completed in 2 minutes 45 seconds, with a perfect score of 5 out of 5, indicating an improvement.

Participant 3 acknowledged during the interview that successfully completing the programming tasks with ChatGPT assistance did not help them learn, but reflecting on and trying to understand the solution during the interview was very helpful. In the interview, Participant 3 commended ChatGPT for its assistance, leading to the learning of new concepts and boosting confidence in programming. Despite acknowledging the potential misuse of AI tools, they expressed eagerness to continue using them for understanding programming concepts better. They did, however, express concerns about potential misuse.

IV. Discussion

The process involved the analyses of data (the scripts from the interviews) and coding, where the codes emerge from the data. The data analysis phase includes repeated comparisons between emerging codes and labels developed. Thus each participant's interview was coded via selective coding independently, applying qualitative verification strategies to compare codes, resulting in the development of thematic labels, followed by a derivation of overarching themes across the dataset.

Highly confident: This student tends to rely heavily on AI tools and is confident in the code generated by the ChatGPT. They are less likely to seek additional support or to verify the accuracy of the code produced.

Cautious: This student is more careful in their use of the ChatGPT and is less confident in the generated code. They are likely to seek additional support and verify the accuracy of the code produced.

Curious: This student is interested in exploring the capabilities of the ChatGPT and is likely to experiment with various prompts to see what types of code the tool generates.

Frustrated: This student struggles with the task and becomes frustrated with ChatGPT. They may give up on using the tool or seek additional support to help them complete the task.

Innovative: This student may use the ChatGPT in creative ways, such as using it to generate code for tasks that are not related to the assignment or to develop new programming concepts.

Further research is needed to examine AI interaction from the perspective of intersectional identities. The perceptions of AI and its association with systemic inequalities in AI literacy levels lead to complex interplays between the student's attitudes towards AI and their behavior when allowed to use AI [15, 16]. The accessibility and availability of AI tools, such as ChatGPT, are not uniformly distributed among different communities and demographic groups [17]. This discrepancy results in unequal AI literacy levels, with certain groups having greater exposure and resources to develop proficiency in AI-assisted programming. These disparities align with broader systemic inequities present in society, perpetuating existing power imbalances [18].

Furthermore, intersectional privileges and barriers further shape individuals' experiences when engaging with AI tools for coding and programming [19, 20]. Variables such as gender, race, socioeconomic status, and disability can significantly influence an individual's access to education, opportunities, and resources related to AI [21, 22, 23]. Consequently, this impacts their ability to effectively leverage AI tools and hampers their potential for growth and success in programming domains [24].

Understanding and addressing these intersectional dynamics is essential to establishing an inclusive and equitable AI ecosystem. Initiatives should be undertaken to bridge the AI literacy gap and provide support and resources to marginalized communities, enabling them to take advantage of the potential of AI tools such as ChatGPT.

V. Conclusion

This paper presents preliminary findings from a study of the use of ChatGPT by computer science students in their second or third semester of learning to program in Java. The study design includes data collection about student identity, previous experience with AI tools, prior knowledge of Java concepts, interaction behavior while solving programming tasks with the help of ChatGPT, and an interview about their perception of AI and their experience using ChatGPT. Our findings show that attitude and experience with AI play an important role in their behavior and use of ChatGPT. These findings provide insight into the planning for the use of ChatGPT to ensure that it benefits all students. Further research is needed to understand the complex relationships between student identity, attitude, and behavior with AI tools such as ChatGPT. With a better understanding of how students use ChatGPT, there is an opportunity to design the interface to better support the student in different types of learning goals: understanding course materials, learning programming concepts, and generating and debugging code.

References

[1] M. Verleger, and J. Pembridge, "A Pilot Study Integrating an AI-driven Chatbot in an Introductory Programming Course," in *2018 IEEE Frontiers Educ. Conf. (Fie)*, Oct. 2018, pp. 1–4, doi: 10.1109/FIE.2018.8659282.

[2] J. Wei, Y. Tay, et al., "Emergent Abilities of Large Language Models," arXiv, 2022. Accessed: Feb. 24, 2023. [Online]. Available: http://arxiv.org/abs/2206.07682

[3] A. Radford, J. Wu, et al., "Language Models are Unsupervised Multitask Learners," 2019. Accessed: Feb. 24, 2023. [Online]. Available: https://cdn.openai.com/better-language-models/language_models_are_unsupervised_multitask_learners.pdf

[4] P. Haluptzok, M. Bowers, and A. T. Kalai, "Language Models Can Teach Themselves to Program Better," arXiv, 2022. Accessed: Feb. 24, 2023. [Online]. Available: http://arxiv.org/abs/2207.14502

[5] P. Vaithilingam, T. Zhang, and E. L. Glassman, "Expectation vs. Experience: Evaluating the Usability of Code Generation Tools Powered by Large Language Models," in *Extended Abstr. 2022 CHI Conf. Human Factors Comput. Syst.* in Chi Ea '22, New York, NY, USA, Apr. 2022, pp. 1–7, doi: 10.1145/3491101.3519665. Accessed: Feb. 24, 2023. [Online]. Available: https://doi.org/10.1145/3491101.3519665

[6] F. Pedró, M. Subosa, A. Rivas, and P. Valverde, "Artificial intelligence in education : challenges and opportunities for sustainable development," 2019. Accessed: Mar. 28, 2023. [Online]. Available: https://unesdoc.unesco.org/ark:/48223/pf0000366994

[7] K. Seo, J. Tang, I. Roll, S. Fels, and D. Yoon, "The impact of artificial intelligence on learner–instructor interaction in online learning," *Int. J. Educational Technol. Higher Educ.*, vol. 18, no. 1, p. 54, Oct. 2021, doi: 10.1186/s41239-021-00292-9. Accessed: May 21, 2023. [Online]. Available: https://doi.org/10.1186/s41239-021-00292-9

[8] M. Wollowski, "Using ChatGPT to produce code for a typical college-level assignment," *AI Mag.*, vol. 44, no. 1, pp. 129–130, 2023, doi: 10.1002/aaai.12086. Accessed: May 21, 2023. [Online]. Available: https://onlinelibrary.wiley.com/doi/abs/10.1002/aaai.12086

[9] M. Wermelinger, "Using GitHub Copilot to Solve Simple Programming Problems," in *Proc. 54th ACM Tech. Symp. Comput. Sci. Educ. V. 1* in Sigcse 2023, New York, NY, USA, Mar. 2023, pp. 172–178, doi: 10.1145/3545945.3569830. Accessed: Apr. 30, 2023. [Online]. Available: https://dl.acm.org/doi/10.1145/3545945.3569830

[10] K. Malinka, M. Perešíni, A. Firc, O. Hujňák, and F. Januš, "On the Educational Impact of ChatGPT: Is Artificial Intelligence Ready to Obtain a University Degree?," arXiv, 2023. Accessed: May 21, 2023. [Online]. Available: http://arxiv.org/abs/2303.11146

[11] C. Geng, Y. Zhang, B. Pientka, and X. Si, "Can ChatGPT Pass An Introductory Level Functional Language Programming Course?," arXiv, 2023. Accessed: May 21, 2023. [Online]. Available: http://arxiv.org/abs/2305.02230

[12] V. Braun, and V. Clarke, "Thematic analysis," in *APA Handbook Res. Methods Psychology, Vol 2: Res. Designs: Quantitative, Qualitative, Neuropsychological, Biol.* (APA Handbooks in Psychology®), Washington, DC, US: American Psychological Association, pp. 57–71.

[13] J. Saldana, "The Coding Manual for Qualitative Researchers," *Coding Manual Qualitative Researchers*, pp. 1–440, 2021. Accessed: May 21, 2023. [Online]. Available: https://www.torrossa.com/en/resources/an/5018667

[14] P. Nasa, R. Jain, and D. Juneja, "Delphi methodology in healthcare research: How to decide its appropriateness," *World J. Methodology*, vol. 11, no. 4, pp. 116–129, Jul. 2021, doi: 10.5662/wjm.v11.i4.116.

[15] A. True-Funk, C. Poleacovschi, et al., "Intersectional Engineers: Diversity of Gender and Race Microaggressions and Their Effects in Engineering Education," *J. Manage. Eng.*, vol. 37, no. 3, p. 4021002, May 2021, doi: 10.1061/(ASCE)ME.1943-5479.0000889. Accessed: May 22, 2023. [Online]. Available: https://ascelibrary.org/doi/10.1061/%28ASCE%29ME.1943-5479.0000889

[16] K. Holstein, and S. Doroudi, "Equity and artificial intelligence in education: will" aied" amplify or alleviate inequities in education?," *Arxiv Preprint Arxiv:2104.12920*, 2021.

[17] T. T. Krupiy, "A vulnerability analysis: theorising the impact of artificial intelligence decision-making processes on individuals, society and human diversity from a social justice perspective," *Comput. Law & Secur. Rev.*, vol. 38, p. 105429, 2020.

[18] Y. K. Dwivedi, N. Kshetri, et al., ""so what if chatgpt wrote it?" multidisciplinary perspectives on opportunities, challenges and implications of generative conversational ai for research, practice and policy," *Int. J. Inf. Manage.*, vol. 71, p. 102642, 2023.

[19] K. Dobransky, and E. Hargittai, "The disability divide in internet access and use," *Information, Communication & Soc.*, vol. 9, no. 3, pp. 313–334, 2006.

[20] S. Ciston, "Imagining intersectional ai," *Xcoax*, p. 39, 2019.

[21] K. Crenshaw, "Demarginalizing the Intersection of Race and Sex: A Black Feminist Critique of Antidiscrimination Doctrine, Feminist Theory and Antiracist Politics," *Univ. Chicago Legal Forum*, vol. 1989, p. 139, 1989. [Online]. Available: https://heinonline.org/HOL/Page?handle=hein.journals/uchclf1989&id=143&div=&collection=

[22] R. S. Baker, and A. Hawn, "Algorithmic bias in education," *Int. J. Artif. Intell. Educ.*, pp. 1–41, 2021.

[23] G. R. Bauer, and D. J. Lizotte, "Artificial intelligence, intersectionality, and the future of public health," vol. 111, American Public Health Association, 2021, pp. 98–100.

[24] N. McDonald, and S. Pan, "Intersectional ai: a study of how information science students think about ethics and their impact," *Proc. ACM Human-Computer Interaction*, vol. 4, no. CSCW2, pp. 1–19, 2020.