

Evaluating Usability and Cognitive Load in Programming Education with ChatGPT Integration

Gustavo Gutiérrez Carreón¹ 

Universidad Michoacana de San Nicolás de Hidalgo, Morelia, Mexico

Rigoberto López Escalera 

Universidad Michoacana de San Nicolás de Hidalgo, Morelia, Mexico

Abstract

This study analyzes the impact of ChatGPT on usability and cognitive load in programming education for undergraduate students in an Information Technology Management program. The goal is to evaluate whether ChatGPT improves learning outcomes in programming topics. A comparative research design was used with two groups: a traditional instruction control group and an experimental group using ChatGPT as a support tool. Data were collected using the System Usability Scale (SUS), a custom usability questionnaire, cognitive load surveys, and programming performance evaluations. The results indicate that the students using ChatGPT reported greater usability and a lower cognitive load. They also showed improved comprehension and problem-solving skills. However, improper use of the tool can lead to superficial learning, highlighting the need for structured guidance. The findings suggest that, when integrated appropriately, ChatGPT can improve the learning experience by reducing mental effort and enhancing participation in programming education. Recommendations are offered to help educators incorporate AI tools effectively and responsibly.

2012 ACM Subject Classification Computing methodologies → Apprenticeship learning

Keywords and phrases Usability, Cognitive Load, ChatGPT-assisted Programming Education

Digital Object Identifier 10.4230/OASICS.ICPEC.2025.12

Funding *Gustavo Gutiérrez Carreón:* CIC-UMSNH.

Rigoberto López Escalera: CIC-UMSNH.

1 Introduction

Artificial intelligence (AI) tools are increasingly integrated into educational environments, reshaping how students learn and interact with content. Among these, large language models (LLMs) such as ChatGPT have emerged as versatile support tools in higher education, capable of generating explanations, helping code, and scaffolding problem solving tasks [8],[1]. In programming education, where abstract reasoning and technical precision are essential, such tools offer potential benefits in reducing mental effort and improving usability.

Programming courses, particularly those covering databases, web development, and cybersecurity, often impose high cognitive demands. According to Cognitive Load Theory, learning effectiveness is influenced by three types of cognitive load: intrinsic, germane, and extraneous [13]. Poorly designed instructional environments can overload learners' working memory, hindering performance and long-term retention. Effective design, therefore, must aim to optimize these load types to promote schema construction and skill development.

ChatGPT may function as a form of intelligent tutoring, supporting real-time learning through interaction. Its ability to provide immediate, context-sensitive assistance positions it as a tool with strong usability potential – defined by effectiveness, efficiency, and user satisfaction [9]. Prior studies have shown that usability and cognitive load are interrelated factors that influence student outcomes in technology-mediated learning [7].

¹ Corresponding author



© Gustavo Gutiérrez Carreón and Rigoberto López Escalera;
licensed under Creative Commons License CC-BY 4.0

6th International Computer Programming Education Conference (ICPEC 2025).

Editors: Ricardo Queirós, Mário Pinto, Filipe Portela, and Alberto Simões; Article No. 12; pp. 12:1–12:7

OpenAccess Series in Informatics



OASICS Schloss Dagstuhl – Leibniz-Zentrum für Informatik, Dagstuhl Publishing, Germany

12:2 Evaluating Usability and Cognitive Load

Despite its growing use, there is limited empirical evidence on how ChatGPT affects students' perceived usability and cognitive load in programming contexts. This study addresses that gap by comparing traditional instruction with ChatGPT-assisted learning among undergraduate students. Using validated instruments, the System Usability Scale (SUS) and a cognitive load questionnaire, along with academic performance measures, the study evaluates ChatGPT's pedagogical impact and discusses implications for instructional design in computing education.

2 Literature Review

The rapid adoption of artificial intelligence (AI) in education has prompted a growing body of research exploring its pedagogical implications. In particular, the emergence of large language models (LLMs) such as ChatGPT has opened new avenues for intelligent tutoring, student engagement, and support in cognitively intensive domains like programming. Studies indicate that these tools can function as on-demand instructional aids, helping learners navigate complex tasks, identify and correct errors, and receive real-time feedback.

One of the central constructs in evaluating AI-based educational tools is usability, commonly defined as the degree to which a system can be used effectively, efficiently, and satisfactorily by its intended users [9]. The System Usability Scale (SUS) developed by [4] and later refined with interpretive benchmarks [2], is widely used to measure this construct. SUS has been validated across diverse digital interfaces, including educational technologies, and is valued for its reliability and simplicity [3].

Closely related to usability is cognitive load, a critical variable in instructional design. According to Cognitive Load Theory (CLT), the effectiveness of a learning experience is determined by the interplay of three types of load: intrinsic, extraneous, and germane [13]. Poor interface design, confusing navigation, or overwhelming instructions can impose extraneous load, which detracts from meaningful learning. Conversely, germane load reflects cognitive resources directed toward schema formation and understanding [14]. Research has shown that usability issues often correlate with increased cognitive load and lower learning performance [10].

In programming education, AI support systems such as ChatGPT are being examined as tools to reduce cognitive load and improve student outcomes. For instance, [15] developed and evaluated an AI tutor for Java programming, reporting improvements in student engagement and task efficiency. Similarly, [7] constructed a synthetic index to measure student-perceived impact of AI tools in higher education, highlighting usability and learning facilitation as key indicators of positive reception.

Despite these promising insights, concerns persist. Some studies warn that unguided use of AI may lead to overreliance or surface learning, where students depend on the tool without fully understanding the underlying concepts [12], [6]. These findings underscore the need for structured integration and instructional mediation when deploying ChatGPT in formal learning environments.

Finally, prior work also emphasizes the relevance of learner experience design (LXD), which combines usability principles with pedagogical goals to shape effective digital learning environments [5], [11]. Incorporating LXD frameworks helps ensure that AI tools like ChatGPT support rather than disrupt the learning process, particularly in cognitively demanding fields such as programming and data science.

In summary, existing literature provides a strong rationale for studying the interplay between ChatGPT, usability, and cognitive load. However, empirical research specifically focused on undergraduate programming education remains limited. This study contributes to closing that gap by systematically evaluating students' experiences with ChatGPT in terms of usability and cognitive load, using established quantitative instruments.

3 Methodology

This study adopted a comparative research design with two groups: an experimental group and a control group. The independent variable was the use of ChatGPT as a support tool in learning Java programming. The dependent variables were the perceived usability of the learning process and the cognitive load experienced by students.

3.1 Participants

Participants were 60 undergraduate students in their fifth semester of the Bachelor's Degree in Information Technology Management at the Faculty of Accounting and Administrative Sciences, Universidad Michoacana de San Nicolás de Hidalgo. Students were divided into two groups:

Experimental Group (n = 30): Used ChatGPT as a support tool during programming tasks.

Control Group (n = 30): Received traditional instruction without ChatGPT.

3.2 Instruments

3.2.1 Usability

Usability was understood as the degree to which a tool allows users to achieve specific learning objectives effectively, efficiently, and satisfactorily. Two instruments were used:

System Usability Scale (SUS): A standardized 10-item Likert-scale questionnaire that provides a global usability score ranging from 0 to 100 [4]. Higher scores indicate better usability.

Custom Usability Questionnaire: Developed specifically for this context, this 5-point Likert-scale instrument evaluated:

- Clarity of ChatGPT's explanations.
- Ease of retrieving relevant information.
- Usefulness in resolving coding errors.
- Overall satisfaction with ChatGPT as a learning aid.

Although SUS was originally designed to evaluate software systems, it has been widely adapted in educational settings to measure perceived ease of use, clarity, and effectiveness of learning tools and environments. In this study, we apply SUS as a proxy for the interactional experience of students with ChatGPT, treating it as a component of the instructional design [14].

3.2.2 Cognitive Load

Cognitive load was measured using a Likert-scale questionnaire targeting:

- Mental effort is required to complete tasks.
- Perceived task difficulty.
- Levels of frustration.
- Uncertainty in problem-solving.

Additionally, programming assessment scores were used as an indirect indicator of cognitive load, based on the assumption that greater difficulty would correlate with lower performance.

3.2.3 Academic Performance

Using the NetBeans development environment, two practical programming assessments were conducted to measure students' comprehension and problem-solving ability in Java.

3.3 Procedure

The study was conducted over a four-week period and followed these steps:

- Common Instruction Phase: All participants received identical theoretical instruction on Java programming fundamentals to ensure baseline consistency.
- Control Group Activity: Students completed programming tasks using only standard resources – NetBeans IDE, course materials, online references, and instructor guidance.
- Experimental Group Activity: In addition to the same tools, students in this group had access to ChatGPT. They were trained to use the model for:
 - Clarifying programming concepts.
 - Generating basic code templates.
 - Receiving suggestions for problem-solving
 - Debugging code with AI assistance

Students were explicitly informed that ChatGPT was a support tool, not a substitute for active learning.

- Assessments: Both groups completed the same two programming evaluations during the study to measure their ability to apply course concepts in solving real-world tasks.
- Post-Intervention Evaluation: At the end of the study, participants completed the SUS, the custom usability questionnaire, and the cognitive load questionnaire.
- Data Collection: Performance scores from the programming evaluations were collected for both groups.

3.4 Data Analysis

Descriptive statistics (mean, standard deviation) were calculated for SUS, cognitive load, and custom usability scores.

Inferential statistics (independent samples t-tests) were used to compare usability, cognitive load, and academic performance between the experimental and control groups.

Thematic analysis was performed on qualitative responses from open-ended questionnaire items and student observations during sessions to extract insights on user experience and perception.

4 Results

As of the submission of this article, data collection for the study has been completed, and preliminary processing of usability and cognitive load responses is underway. Preliminary SUS scores in the experimental group ($M = 78.5$, $SD = 10.2$) were higher than in the control group ($M = 65.4$, $SD = 12.6$), suggesting improved perceived usability with ChatGPT support. However, a full statistical analysis of the results is still in progress. Therefore, the findings presented in this section are partial and intended to outline the ongoing research efforts rather than to offer definitive conclusions.

Initial observations suggest that students in the experimental group – those who used ChatGPT as a support tool – reported higher perceived usability, as indicated by early trends in System Usability Scale (SUS) scores. Similarly, responses to the custom usability questionnaire indicate that students valued ChatGPT's clarity in explanations and its usefulness in resolving programming errors.

Regarding cognitive load, preliminary descriptive statistics show lower self-reported levels of extraneous load among the experimental group. Students commonly reported that ChatGPT helped reduce frustration and uncertainty when encountering programming challenges. These early patterns are consistent with existing literature on AI-supported learning environments [15].

However, the complete statistical analysis – including t-tests to compare usability and cognitive load scores between groups, as well as performance differentials on programming evaluations – remains to be finalized. The qualitative data, consisting of open-ended survey responses and observational notes, is also currently being coded thematically to identify patterns in student experience and tool usage.

Future work will involve:

- Completing the full quantitative analysis of all collected data
- Cross-validating cognitive load findings with academic performance
- Integrating qualitative themes to better understand learner behavior with ChatGPT

These results will be critical in refining instructional strategies for AI-supported programming education and in establishing best practices for integrating LLMs into the curriculum.

5 Discussion

The preliminary findings of this study suggest that the integration of ChatGPT into programming instruction may offer notable benefits in terms of usability and cognitive support. Although the final statistical analysis is still in progress, early data trends indicate that students perceived the AI tool as a helpful complement to traditional instruction, particularly in clarifying programming concepts and reducing frustration during problem-solving.

These observations align with prior research that highlights the potential of AI tutors to enhance user experience and task efficiency in technical disciplines [15]. The improved usability scores observed in the experimental group reinforce the idea that tools like ChatGPT can make learning environments more accessible and responsive to students' immediate needs [9].

From a cognitive load perspective, the observed reduction in extraneous load supports the theoretical assumption that well-designed technological scaffolds can alleviate mental burden during complex tasks [14]. By offloading lower-level struggles – such as syntax errors or debugging assistance – students may have more cognitive capacity available for deeper conceptual engagement.

However, these early benefits must be interpreted with caution. Literature has warned about the risks of overdependence on AI systems, which may lead to superficial learning or passive consumption of content ([12]. In this study, students were explicitly guided to use ChatGPT as a support tool, not as a replacement for active learning – a distinction that is essential for maintaining educational integrity.

Another important consideration is that perceived usability and reduced cognitive load do not automatically translate into improved academic performance. While initial feedback suggests a positive learner experience, the correlation between these perceptions and actual learning outcomes remains to be confirmed through final performance analysis.

This study, therefore, contributes to the ongoing discourse around learning experience design (LXD) and its role in shaping AI-enhanced educational environments [11]. By incorporating usability, cognitive psychology, and human-computer interaction frameworks, this research builds a multidisciplinary foundation for the responsible integration of LLMs into programming education.

In summary, while early results are promising, comprehensive data analysis is required to validate the impact of ChatGPT on learning effectiveness. Future studies should explore long-term effects, student autonomy, and the role of instructional mediation to ensure AI tools enhance rather than replace essential cognitive processes in programming education.

6 Conclusions and Future Work

This study explores the impact of integrating ChatGPT as a support tool in undergraduate programming education, focusing on perceived usability and cognitive load. Preliminary findings suggest that students who used ChatGPT reported higher usability and lower extraneous cognitive load compared to those receiving traditional instruction. These early results are consistent with existing literature that highlights the potential of AI-powered assistants to enhance learner experience, reduce cognitive strain, and support self-directed learning.

The study contributes to a growing body of research that positions large language models not merely as informational tools but as dynamic learning aids capable of adapting to students' needs in real time. By applying validated instruments such as the System Usability Scale (SUS) and cognitive load questionnaires, this research offers a structured framework for evaluating how generative AI can be responsibly integrated into programming instruction.

However, as the study is still in progress, the full analysis of the quantitative and qualitative data remains ongoing. Key next steps include completing statistical comparisons between groups, analyzing academic performance outcomes, and conducting thematic coding of qualitative feedback. These analyses will provide a more definitive assessment of ChatGPT's pedagogical value and its implications for instructional design.

For future research, several directions are proposed:

- Longitudinal studies to assess the sustained impact of AI tools on learning and retention.
- Controlled experiments with larger and more diverse student populations to increase generalizability.
- Investigations into learning strategies used by students when interacting with AI, particularly in how they balance support with independent problem-solving.
- Design guidelines for educators on how to integrate generative AI responsibly and effectively into programming curricula.

Ultimately, the findings from this and future studies aim to inform evidence-based practices for using AI in education – ensuring that technological innovation is aligned with cognitive principles and pedagogical goals.

References

- 1 Ibrahim Adeshola and Adeola Praise Adepoju. The opportunities and challenges of chatgpt in education. *Interactive Learning Environments*, 32(10):6159–6172, 2024.
- 2 Aaron Bangor, Philip Kortum, and James Miller. Determining what individual sus scores mean: Adding an adjective rating scale. *Journal of usability studies*, 4(3):114–123, 2009.
- 3 Aaron Bangor, Philip T Kortum, and James T Miller. An empirical evaluation of the system usability scale. *Intl. Journal of Human-Computer Interaction*, 24(6):574–594, 2008.

- 4 John Brooke et al. Sus-a quick and dirty usability scale. *Usability evaluation in industry*, 189(194):4–7, 1996.
- 5 Ruth C Clark, Frank Nguyen, and John Sweller. *Efficiency in learning: Evidence-based guidelines to manage cognitive load*. John Wiley & Sons, 2011.
- 6 Laode Muhamad Fathun, Chiara Vincha, Aisha Saharani, Algis Zalita Pitra, et al. Knowledge sharing implications of the emergence of gpt chat on learning methods at smp 4 muhammadiyah depok, west java. *Jurnal Pengabdian dan Pemberdayaan Masyarakat Indonesia*, 3(3):130–138, 2023.
- 7 Alberto Grájeda, Johnny Burgos, Pamela Córdova, and Alberto Sanjinés. Assessing student-perceived impact of using artificial intelligence tools: Construction of a synthetic index of application in higher education. *Cogent Education*, 11(1):2287917, 2024.
- 8 Ahmad Haidar. Chatgpt and generative ai in educational ecosystems: Transforming student engagement and ensuring digital safety. In *Preparing Students for the Future Educational Paradigm*, pages 70–100. IGI Global Scientific Publishing, 2024.
- 9 James R Lewis and Jeff Sauro. Item benchmarks for the system usability scale. *Journal of Usability studies*, 13(3), 2018.
- 10 Elena Novak, Jerry Daday, and Kerrie McDaniel. Assessing intrinsic and extraneous cognitive complexity of e-textbook learning. *Interacting with Computers*, 30(2):150–161, 2018.
- 11 Matthew Schmidt and Rui Huang. Defining learning experience design: Voices from the field of learning design & technology. *TechTrends*, 66(2):141–158, 2022.
- 12 Muhammad Shidiq. The use of artificial intelligence-based chat-gpt and its challenges for the world of education; from the viewpoint of the development of creative writing skills. In *Proceeding of international conference on education, society and humanity*, volume 1, pages 353–357, 2023.
- 13 John Sweller. Cognitive load theory and educational technology. *Educational technology research and development*, 68(1):1–16, 2020.
- 14 Andrew A Tawfik, Linda Payne, Andrew M Olney, and Heather Ketter. Exploring the relationship between usability and cognitive load in data science education. *The Journal of Applied Instructional Design*, 12(3), 2023.
- 15 Alessia Tripaldelli, Brian Butka, and Casey Elder. The development and evaluation of artificial intelligence (ai) tutor for a java programming class.