EVALUATING CHATGPT'S EFFECTIVENESS IN FORMATIVE ASSESSMENT PRACTICES FOR CS EDUCATION: A SWOT ANALYSIS

Jacqui Chetty
University of Birmingham
United Kingdom

Abstract

Recently, large language models (LLMs) such as OpenAI's ChatGPT and AI-powered IDEs like GitHub Copilot have become increasingly integrated into teaching and learning. This study evaluates the effectiveness of ChatGPT for formative assessment through a Strengths, Weaknesses, Opportunities, and Threats (SWOT) analysis. The SWOT analysis serves as an evaluative tool to critically assess the effectiveness of ChatGPT in assessment practices. The findings suggest that ChatGPT is effective as an automated tutor; supports students in engaging with programming concepts; provides a real-time environment for students to rapidly assess and reflect on their code; and enhances their learning experience. However, caution is necessary, as ChatGPT may produce inaccuracies, lack deep conceptual awareness, and pose risks to academic integrity.

Introduction

In late 2022 OpenAI launched ChatGPT, seen as a breakthrough Large Language Model (LLM) that could generate text and maintain human-like conversations (Rahman & Watanobe, 2023). ChatGPT and similar LLMs have the potential to create opportunities, present challenges, pose threats, raise ethical concerns and disrupt education fields. For example, for university students, ChatGPT can assist in research and writing tasks and develop critical thinking and problem solving (Kasneci et al., 2023). It can facilitate group and remote learning and empower learners with disabilities by combining speak-to-text or text-to-speech solutions. For educators, ChatGPT can assist with personalising student learning, lesson planning, research, and writing as well as assessment and evaluation. However, the negative impacts of easy cheating and plagiarism, ChatGPT solving problems instead of students acquiring a skill set, and providing incorrect knowledge (Malinka et al., 2023), can outweigh the advantages.

Although a very new research field, many researchers are investigating the use of LLMs in education. The primary focus is on improving the learning process and aiding students as well as re-designing repetitive processes for educators. For

example, Qureshi (2023) explored the use of ChatGPT as a tool for learning and assessment in undergraduate computer science, highlighting the opportunities and challenges. This would be no different for higher education as ChatGPT is seen as a potential disruptor to teaching-and-learning. Banerjee et al. (2025) conducted an impact analysis by evaluating the capability of ChatGPT for instructional purposes in the field of computer science and engineering. The article explores the opportunities and limitations of ChatGPT as well as performing a student survey to highlight ChatGPT anomalies and concerns. Tlili et al. (2023) discuss whether ChatGPT is friend or foe, including the extent to which ChatGPT has additional qualities, such as personality, emotion, its usefulness, ethical considerations, cheating, and truthfulness. The article highlights that to incorporate ChatGPT into instructional design, the way knowledge is assessed must change.

Much of the research focus thus far is aimed at strengths, weaknesses and threats. Accordingly, conducting a SWOT analysis (Strengths, Weaknesses, Opportunities, and Threats) is a logical course of action to evaluate ChatGPT in large-scale assessment practices. The SWOT would highlight ChatGPT's potential benefits, limitations, and future implications for educational settings. The SWOT analysis is framed as an evaluative framework to critically examine ChatGPT's effectiveness in assessment practices.

The aim of the paper is to address the following research questions:

> **RQ1:** What are the strengths and weaknesses of ChatGPT as an educational tool for formative assessment in CS education?
> **RQ2:** What external opportunities and threats influence the effectiveness of ChatGPT in formative assessment for CS education?

To answer these research questions a SWOT Analysis was conducted.

Background Motivation

Existing Assessment Practices

Covid-19 changed teaching-and-learning and assessment practices dramatically world-wide. Overnight, alternatives to traditional ways of teaching were adopted and e-learning was the best option available to ensure learning continued and students and educators were kept safe (Maatuk et al., 2021). The UK adopted a similar approach where teaching was online, and assessments were "take home" assignments. Students completed these, uploading the completed assessment to a learning management system (LMS). Post-Covid, traditional assessment practices (exam-styled) have been re-introduced, in combination with online assessments that are completed remotely, uploaded to an LMS.

Current Challenges

During the Covid era university enrolment increased dramatically. Post Covid, these numbers have remained high, placing pressure on educators and students alike. Large cohorts enrolled in a course means that educators are under pressure to provide quality learning, assessment practices, feedback, and support, to large numbers of students.

With the rise of publicly accessible large language models (LLMs) such as ChatGPT and less one-to-one access to educators, students are increasingly relying on these tools when learning programming. This can potentially be detrimental when learning a skill like programming as coding requires practice. The use of ChatGPT may interfere with the learning process. However, ChatGPT offers an appealing shortcut, particularly for beginners.

Current challenges faced by students and educators are:

- Students struggle to assess their performance (Tam, 2021).
- Post Covid, students struggle to transition from "take home" assessments to exam-style ones (Aboagye et al., 2020).
- The continued reliance on ChatGPT may contribute to students not developing the essential problem-solving and coding skills required (Hermans, 2021).
- The reliance on ChatGPT does not guarantee enhanced performance (Becker et al., 2023; Xue et al., 2024).
- Time and attention given to smaller groups of students is reduced.
- Assessments completed as "take home" or online, over an extended period, can lead to students approaching programming tasks with the support of other students and the use of ChatGPT.
- Educators are unable to assess students' abilities when assessments are completed online.
- ChatGPT raise concerns about academic integrity and fairness.

Research shows that ChatGPT may be a good tool to use as part of a layered approach to learning (Becker et al., 2023). For example, students can verify their code using ChatGPT and constructive feedback can further their learning so ChatGPT can then be used towards formative assessment.

Methodology

This study employs a SWOT analysis framework to evaluate ChatGPT's effectiveness for formative assessment using a quantitative research approach.

Description of the Module

For the 2024/25 academic year the cohort (n=318) registered for a continuous assessment (CA) module, aimed at master's students learning programming for the first time. Students are either completing a year in computer science (YiCS) or registered for a degree conversion. Sixty percent of students have little to no programming experience. Within one term (12 weeks) students learn and engage with Python, used as a vehicle, to teach the fundamental programming concepts of programming, as well as more advanced concepts such as object-oriented programming and inheritance.

Formative Assessment Opportunities

Weekly formative assessment takes place in a lab. Students are tasked with completing worksheets, based on the content taught to them for that week. They are then asked to complete a survey, shown in Table 1, regarding the learning and the worksheet completed. Surveys are only conducted for the first half of the term, due to survey fatigue.

Table 1

Weekly Surveys Regarding Labs

Question	***Multichoice***	***Week 2 (n=88)***	***Week 3 (n=41)***	***Week 4 (n=29)***
I was able to comfortably master the concepts taught to me in the lecture this week.	*Absolutely*	*22 (25%)*	*5 (12%)*	*5 (17%)*
	Somewhat	*50 (57%)*	*30 (73%)*	*21 (72%)*
	Not really	*16 (18%)*	*6 (15%)*	*3 (10%)*
I managed to complete most of the exercises on the worksheet this week.	*Absolutely*	*59 (67%)*	*14 (34%)*	*18 (62%)*
	Somewhat	*28 (32%)*	*27 (66%)*	*11 (38%)*
	Not really	*1 (1%)*	*0 (0%)*	*0 (0%)*

Additionally, Table 2 shows the topics that students found difficult.

Table 2

Difficulty Learning Fundamental Programming Concepts

Week 2		**Week 3**		**Week 4**	
Concept	*Number reporting difficulty*	*Concept*	*Number reporting difficulty*	*Concept*	*Number reporting difficulty*
Calling functions	13	If…else	2	I still struggle with loops	6
Writing functions	17	While loops	12	Lists	2
Function signatures	22	For loops	11	Dictionaries	8
*args	44	Lists	48	Understanding larger applications	18
*kwargs	59				
Total	**155**		**73**		**34**

The challenge within this educational setting is the teaching assistant (TA) to student ratio, which averages 1:35. To alleviate the pressure on TAs, the use of ChatGPT was integrated into the lab for weeks 3 and 4 (see Table 3). The aim was for students to solve a problem and develop a solution in Python. They were instructed to engage with ChatGPT (as opposed to the TA's) to assist them with any difficulties they encountered when coding. As seen in Table 3, for week 3, students struggled to implement their own solution, however by week 4, there struggle was less notable. Table 3 also shows that students engaged with ChatGPT to explore incorrect coding; when they had difficulty creating a solution; and when they were grappling to formulate a solution. Further investigation is required to determine if the explanations provided by ChatGPT bridged the gap.

In week 9 students were tasked with completing a programming problem. They were also instructed to ask ChatGPT for a solution to the same problem and compare their solution to that of ChatGPT. Finally, they completed a quiz where the following question was put to them:

> "I have completed the solution as instructed in exercise 3 for this week's lab worksheet. I have studied the feedback and alternatives given to me by ChatGPT. I feel that I should be awarded the following grade (0 to 10) for my solution".

Table 3

Weekly Surveys Regarding ChatGPT Experience

Question	***Multichoice***	***Week 3***	***Week 4***
The solution that I implemented was like that of ChatGPT	*Yes*	*16*	*23*
	No	*19*	*13*
	I struggled to implement a solution	*25*	*5*
After reading ChatGPT's explanation of the code can you describe the code to someone	*Yes*	*49*	*39*
	No	*11*	*2*
After ChatGPT explained the code, I better understood how to solve the problem	*Absolutely*	*33*	*26*
	Somewhat	*25*	*14*
	Not really	*2*	*7*
Using ChatGPT helps me learn how to code	*Absolutely*	*38*	*27*
	Somewhat	*21*	*14*
	Not really	*1*	*0*
After learning how to create this function with the help of ChatGPT, could you now code something similar on your own?	*Yes*	*49*	*24*
	No	*11*	*16*
Do you trust that ChatGPT is providing you with correct knowledge	*Yes*	*38*	*30*
	No	*21*	*11*

Table 4 shows the outcome of the number of students that allocated a grade to themselves based on their solution and comparing it to ChatGPT's solution. Although the results show that the average grade was 8.9, it does seem that students graded themselves towards the higher end. However, when reviewing the reasons for the grades, students graded themselves very fairly and provided valid reasons for the grade. In many cases they often grading themselves downwards.

Table 4

Student Self-evaluation of Their Programming Solution (0 – 10)

0	***1***	***2***	***3***	***4***	***5***	***6***	***7***	***8***	***9***	***10***
2	*0*	*0*	*4*	*1*	*2*	*12*	*15*	*64*	*66*	*53*

Summative Assessment Opportunities

Although not part of this study, it is important to note that summative assessments were also conducted. These consisted of two proctored (invigilated) two-hour assessments. For the proctored assessments, students are allowed access to learning materials. Additionally, students are allowed access to selected educational sites; however, access to ChatGPT is not allowed. During proctored tests, invigilators and plagiarism detection systems ensure compliance. Prior to the proctored assessments, students are encouraged to complete mock tests. They are expected to complete these mocks independently; however, support is provided. For all proctored tests, to ensure integrity, multiple monitoring systems are employed:

- Extended time and small rooms are provided to students with reasonable adjustment plans (RAPs).
- Invigilators supervise the timed tests.
- Manual grading enables instructors to identify ChatGPT-generated solutions (often abstract or generic and misaligned with coding techniques emphasised in the curriculum).
- Automated plagiarism detection tools compare student submissions.

The grading is managed by six educators, each having access to a shared spreadsheet that finely details how scores are allocated. Additionally, a column for feedback is also included. The grade and the feedback are provided to students. Table 5 shows the overall performance of the cohort comparing 2023/24 (inflated grades due to the Covid era of "take home" assessments) and 2024/25 ("take home" assessments were replaced with proctored assessments to ensure academic fairness).

Table 5

Performance for the Module over 2 Years

Academic year	***Students (n)***	***Pass rate (%)***	***Failures (n, %)***	***t-test: Pass Rate (p-value)***
2024/25	*318*	*67%*	*74 (23%)*	*p = 0.015 * (2023/24 – 2024/25)*
2023/24	*187*	*75%*	*5 (3%)*	*p = 0.042 * (2022/23 – 2023/24)*

Formative Assessment: SWOT Analysis

A SWOT is a structured planning tool used in research (and other sectors) to evaluate internal and external factors regarding a topic. As a research methodology SWOT can assist in assessing the effectiveness regarding an area of interest.

To answer the two research questions, the data from the formative assessments was analysed using a SWOT analysis.

Results and Discussion

Table 6 presents the results of the SWOT analysis derived from the quantitative data in Tables 2, 3 and 4.

Table 6

SWOT Analysis to Answer RQ1 & RQ2

Strengths	Weaknesses	Opportunities	Threats
1. Effective learning support 2. Encourages independent problem-solving 3. Alleviates teaching assistant (TA) workload 4. Self-evaluation and reflection	1. Potential for inaccuracy and misconceptions 2. Struggles with implementation 3. Limited personalisation and adaptive feedback	1. Enhancing automated tutoring capabilities 2. Developing AI literacy among students 3. Bridging the TA gap in large cohorts 4. Gamification and interactive learning	1. Academic integrity concerns 2. Varying accuracy and bias in AI responses 3. Resistance to AI adoption

To provide a more comprehensive explanation of the SWOT analysis mapped in Table 6, further discussion is required to reflect on the pedagogical, technical and ethical considerations when contemplating ChatGPT as an educational tool to support formative assessments.

Strengths

1. Effective learning support
   - Most students found that ChatGPT helped them to code (Week 3, Week 4).
   - Students reported improved understanding after engaging with ChatGPT (Week 3, Week 4).
2. Encourages independent problem-solving
   - Many students felt confident in coding similar problems after using ChatGPT (Week 3, Week 4).
   - ChatGPT explanations improved students' ability to describe code (Week 3, Week 4).
3. Alleviates TA workload
   - Given the high TA-to-student ratio (1:35), ChatGPT serves as an additional learning resource, reducing reliance on human assistance (Week 3).
4. Self-evaluation and reflection
   - Students compared their solutions with ChatGPT (Week 9), developing self-assessment skills (Table 4).

Weaknesses

1. Potential for inaccuracy and misconceptions
   - Some students did not trust ChatGPT's responses (week3, Week 4).
   - ChatGPT lacks deep conceptual awareness as students still struggled with function parameters and loops (seen in Table 2).
2. Struggles with implementation
   - Some students struggled to implement solutions even after using ChatGPT (Week 3, Week4).
3. Limited personalisation and adaptive feedback
   - Unlike human instructors, ChatGPT does not adapt explanations to individual student needs in real-time (Week 4).

Opportunities

1. Enhancing automated tutoring capabilities
   - Integrating ChatGPT with personalised hints and adaptive scaffolding could improve effectiveness.
2. Developing AI literacy among students
   - Using ChatGPT helps students critical analyse AI-generated content, fostering AI literacy and debugging skills.
3. Bridging the TA gap in large cohorts

- Expanding ChatGPT's role in routine formative assessments can further support students without additional staffing costs.

4. Gamification and interactive learning
   - Incorporating ChatGPT into gamified coding challenges could make learning more engaging.

Threats

1. Academic integrity concerns
   - Students may become over-reliant on AI-generated solutions impacting original problem-solving skills.
2. Varying accuracy and bias in AI responses
   - ChatGPT-generated code may contain errors or inefficiencies potentially reinforcing misconceptions (Week 3, Week 4).
3. Resistance to AI adoption
   - Some educators and students may distrust or resist using AI in assessment and learning (Week 3, Week 4).

Summary of SWOT Analysis

In relation to the research questions, the SWOT analysis underscores that ChatGPT offers notable educational benefits for supporting formative assessment in CS education. However, it also draws attention to potential challenges and risks associated with its use. ChatGPT can function as a valuable personal tutor, fostering independent learning. At the same time, over-reliance on such tools may impede students' conceptual understanding and problem-solving development. While ChatGPT has the potential to enhance problem solving abilities, it may inadvertently hinder learning if students adopt inefficient or incorrect coding solutions without critically evaluating the code. Thus, ChatGPT presents a double-edged sword: its effectiveness is highly dependent on how it is integrated into pedagogical practice.

Conclusion

The SWOT analysis highlights ChatGPT's potential to serve as an automated tutor and a valuable educational tool in supporting students with formative assessments. However, its integration must be approached with caution, given notable limitations such as concerns around academic integrity, ethical implications, accuracy, over-reliance, and constraints in delivering personalised feedback.

Future work will focus on the development of robust AI-assisted learning frameworks; the enhancement of critical AI literacy amongst students and

educators; and proactive measures to address ethical considerations related to academic integrity.

References

Aboagye, E., Yawson, J. A., & Appiah, K. N. (2020). COVID-19 and E-Learning: The challenges of students in tertiary institutions. *Social Education Research, 2*(1), 1-8. https://doi.org/10.37256/ser.212021422

Banerjee, P., Srivastava, A. K., Adjeroh, D. A., Reddy, R, & Karimian, N. (2025). Understanding ChatGPT: Impact analysis and path forward for teaching computer science and engineering. *IEEE Access, 13*, 11049-11069. https://ieeexplore.ieee.org/abstract/document/10833612

Becker, B. A., Craig, M., Denny, P., Keuning, H., Kiesler, N., Leinonen, J., Luxton-Reilly, A., Malmi, L., Prather, J., & Quille, K. (2023). Generative AI in Introductory Programming. *Computer Science Curricula*, 1-25.

Hermans, F. (2021). *The Programmer's Brain: What every programmer needs to know about cognition*. New York USA: Manning Publications Co.

Kasneci, E., Sessler, K., Küchemann, S., Bannert, M., Dementieva, D., Fischer, F., Gasser, U., Groh, G., Günneman, S., Hüllermeier, E., Krusche, S., Kutyniok, G., Michaeli, T., Nerdel, C., Pfeffer, J., Poquet, O., Sailer, M., Schmidt, A., Seidel, T.,. . . Kasneci, G. (2023). ChatGPT for good? On opportunities and challenges of large language models for education. Learning and Individual Differences, 103, 102274. https://doi.org/10.1016/j.lindif.2023.102274

Maatuk, A. M., Elberkawi, E. K., Aljawarneh, S., Rashaideh, H., & Alharbi, H. (2021). The COVID-19 pandemic and E-learning: challenges and opportunities from the perspective of students and instructor. *Journal of Computing in Higher Education, 34*, 21-38. https://doi.org/10.1007/s12528-021-09274-2

Malinka, K., Peresíni, M., Firc, A., Hujnák, O., & Janus, F. (2023). On the educational iImpact of ChatGPT: Is artificial intelligence ready to obtain a university degree? *ITiCSE 2023: Proceedings of the 2023 Conference on Innovation and Technology in Computer Science Education V. 1 (pp.* 47 - 53). https://doi.org/10.1145/3587102.3588827

Qureshi, B. (2023). (2023). Exploring the use of ChatGPT as a tool for learning and assessment in undergraduate computer science curriculum: Opportunities and challenges. *arXiv:2304.11214*. doi:http://dx.doi.org/10.48550/arXiv.2304.11214

Rahman, Md. M., & Watanobe, Y. (2023). ChatGPT for education and research: Opportunities, threats, and strategies. Applied Science, 13(9), 5783. https://doi.org/10.3390/app13095783

Tam, A. C. F. (2021). Students' perceptions of and learning practices in online timed take-home examinations during Covid-19. Assessment & Evaluation in Higher Education, 47(3), 477-492. https://doi.org/10.1080/02602938.2021.1928599

Tlili, A., Shehata, B., Adarkwah, M. A., Bozkurt, A., Hickey, D. T., Huang, R., & Agyemang, B. (2023). What if the devil is my guardian angel: ChatGPT as a case study of using chatbots in education. Smart Learning Environments, 10, 15. https://doi.org/10.1186/s40561-023-00237-x

Xue, Y., Chen, H., Bai, G. R., Tairas, R., & Huang, Y. (2024). Does ChatGPT help with introductory programming? An experiment of students using ChatGPT in CS1. *ICSE-SEET '24: Proceedings of the 46th International Conference on Software Engineering: Software Engineering Education and Training* (pp. 331 - 341). https://doi.org/10.1145/3639474.3640076

Author Details

Jacqui Chetty
School of Computer Science, College of Engineering and Physical Sciences
University of Birmingham
UNITED KINGDOM
j.chetty@bham.ac.uk