Students' perceptions of an AI Chatbot with integrated course material in introductory programming

Rolando Gonzalez[1], Toktam Ramezanifarkhani[1], Sturla Bakke[1], Gro Skåland[1], and Morten Forsberg[1]

Kristiania University of Applied Sciences, Oslo, Norway

Abstract. AI chatbots are now widely available, and educational institutions are exploring how students can use them to support their learning, as well as the pedagogical issues arising from that use. In this study we explore how a customized AI chatbot that makes use of gpt-4o, and has access to the course material, is perceived by students in an introductory programming course in a higher education institution, through the lens of Activity Theory and Vygotsky's zone of proximal development, asking to what extent a course-bound chatbot can take on the role of a more knowledgeable other. Data was collected at three points in 2025 using semi-structured interviews: an initial test with 4 student assistants, interviews with 2 students taking the course, and three focus groups with a total of 9 students taking the course. Our findings indicate that the AI chatbot works well for explaining theory, helps identifying errors, and especially can provide step-by-step guides which may lead to a dialogic-style learning process. However, if the chatbot isn't configured right, or the students don't include the right instructions in the prompt, it may end up solving instead of helping with the tasks, resulting in a lesser learning experience. We also find that the perceived warmth of the AI chatbot can be of importance for the students' motivation and adoption.

Keywords: AI Chatbot · Introductory programming· gpt-4o · ZPD

1 Introduction

As Artificial Intelligence (AI) chatbots have become more available and sophisticated, we as educational institutions explore new ways to integrate them into our education to improve our students' learning. Every student with an AI chatbot available can now use it as a personal tutor, available around the clock. For students learning programming, it may give step-by-step guidance on how to solve problems, alternative solutions to tasks, and help with finding errors.

This study explores how students perceive benefits and issues arising from the use of a customized chatbot built on gpt-4o that has access to the syllabus. Chapter 4 presents the students' perceptions across three collected datasets.

2 Related literature

2.1 Activity Theory and the Zone of Proximal Development.

The Zone of Proximal Development complements Activity Theory by offering a more detailed view of how cognitive development unfolds through social interactions. Introduced by Lev Vygotsky, the Zone of Proximal Development is defined as the difference between what a learner can achieve independently and what can be accomplished with assistance from someone more knowledgeable, termed the "More Knowledgeable Other" (MKO). The ZPD represents the competence level slightly beyond the learner's current capabilities, the zone of current development, that becomes attainable through guided support [22].

Activity theory emphasises that learning takes place in a social and cultural context in which tools (including language and technology) and interaction play a central role. Stojanov [20] points out that AI technology such as ChatGPT acts as a 'knowledgeable other', using language and dialogue as tools for learning in a scaffolding manner, and that the "scaffolder" need not be a single expert: in classrooms it is distributed among fellow students and artefacts, and much of it happens in talk. Mercer's notion of *exploratory talk* describes reasoning made visible and thereby available for challenge among participants [15]. Dialogic teaching builds on this tradition and has been shown to support students' agency in technology-rich settings [16]. It is the tradition from which the system message of our chatbot was written (see chapter 3.2). In our context, we can view the student's interaction with the AI tutor as work towards the student's goal of mastering basic programming skills.

AI chatbots as a More Knowledgeable Other. The claim that generative AI can occupy the MKO position is now made frequently, and with some confidence: an AI tutor may in principle infer the learner's current capabilities and offer timely, context-sensitive assistance, extending Vygotsky's MKO with support that is personalised, scalable and consistently available. Tran et al. [21] argue that a generative model scaffolds through argumentative dialogue, personalises to the individual learner's gaps, and can draw on more knowledge than any single teacher could possess. Cai et al. [4], reviewing the integration of AI tools in higher education through a ZPD lens, arrive at broadly similar conclusions. Ferguson et al. [9] offer a strong empirical version of the argument: instructions personalised in real time by an AI system reduced learners' cognitive load and improved learning outcomes relative to static instructions, which is the contingency requirement being satisfied by machine. In our case, the AI chatbot is the intended MKO, and optimal learning then depends on balancing the challenge within the zone, neither too easy nor too difficult [9].

There are, however, problematic aspects in utilising an AI chatbot as the *more knowledgeable other*. One is that a language model has no independent access to the learner's *current* zone of development; it infers this from the prompt, and a prompt asking for a solution is read as a request from someone who *wants* a solution, rather than from someone who wants to be scaffolded.

2.2 Benefits and issues of using AI chatbots for learning.

Yilmaz and Yilmaz [25] identify several advantages of using ChatGPT for learning programming. It is accessed through natural-language interaction, making it relatively easy for students to formulate prompts, and it can provide explanations and code examples that may help bridge the gap between theoretical concepts and programming. The conversational format also supports iterative learning, as one may ask follow-up questions, request alternative explanations, and progressively refine one's understanding. They further report that ChatGPT can assist with debugging, a challenging activity for novice programmers, and that immediate feedback reduces the time students spend searching for information. In a later study they found that the use of ChatGPT significantly improved students' programming self-efficacy and motivation [26]. Groothuijsen et al. [12] similarly identified benefits including debugging, getting explanations, improving conceptual understanding, and generating and optimizing code. Two systematic reviews, of 67 and 57 studies respectively, identify immediate access to explanations, feedback and learning support, and report that students use ChatGPT for personalized learning, programming, debugging and collaboration [13, 1]. A meta-analysis of 69 experimental studies found improvements in academic performance, affective-motivational outcomes and higher-order thinking, alongside reduced mental effort [6], and one on programming education found a small-to-medium effect on post-test performance and a medium-to-large one on practice performance [5].

The same sources report corresponding difficulties. Students may need guidance, as AI chatbots do not necessarily structure the learning process; some perceive a risk of laziness, as their reliance on the chatbot may reduce the willingness to solve problems themselves. Some experience anxiety because the chatbot does not always provide correct answers [25]. This matters in programming education, where an apparently plausible explanation or piece of code may contain subtle errors, and Yilmaz and Yilmaz therefore emphasize the importance of developing prompting skills and critically evaluating responses. Silva et al. [18] found a large majority of students considered ChatGPT a valuable support tool, used to strengthen comprehension, clarify doubts, obtain hints, identify errors and obtain code examples; but only 8.1 percent agreed that AI could entirely replace teachers in programming, and students were concerned with over-reliance on AI-generated code and with not developing fundamental programming skills. Groothuijsen et al. [12] identified concerns among both teachers and students, including declining code quality, reduced learning, and a potential negative influence on collaboration, including pair programming. Sonlu et al. [19] found cases in which chatbot responses were harder to understand than expected from a human teacher, who may adapt an explanation to the student's level, and some respondents found the answers too long. The distinction between providing information and providing effective pedagogical support matters: a technically comprehensive answer may not be aligned with the learner's needs. Reviews likewise identify over-reliance, reduced independent learning and academic integrity concerns [13, 1], and Deng et al. [6] emphasize that improvements in the quality

of students' outputs do not necessarily demonstrate that students have acquired skills. The educational value of AI chatbots therefore appears to depend on how students use them: used to get answers it may reduce opportunities for independent problem solving, and used to request explanations, feedback and alternative solutions it affords a more constructive learning experience.

2.3 The AI's personality and its effects on Learning.

One primary objective of using AI chatbots is to facilitate self-directed learning [17, 19], where students regulate their own learning process. In such interactions the chatbot is not merely a source of information; its conversational style may also influence how students perceive, trust and engage with the system. Lan et al. [14] report that emotional experiences become part of interactions with AI chatbots and influence usage patterns, that adoption, trust and attachment are shaped by anthropomorphism, responsiveness and perceived warmth, and that students may perceive a chatbot not just as a tool but as an interactive partner.

Richter [17] examines chatbots along the dimensions of warmth, meaning friendliness and approachability, and competence, meaning the perceived ability to give accurate and useful assistance. Both influence how students perceive the chatbot's support for self-directed learning, so personality is not only cosmetic; it can affect students' willingness to use the tool. Sonlu et al. [19] and Richter both report that stronger personality traits can increase engagement motivation and improve learning outcomes, and that motivational language encourages continued reflection, with students generally preferring a warmer chatbot that gave comprehensive answers and asked follow-up questions. A warmer personality is not preferable in all situations, however. Sonlu et al. found a reserved, formal demeanor less engaging, whereas Richter reports that some students preferred it because it was more direct and better for goal-oriented learning. This trade-off between warmth and efficiency suggests that an effective educational chatbot should balance warmth with competence rather than maximize either in isolation. Letting students customize conversational style could be one answer, though personalization needs care, since emotional attachment may make students trust the chatbot without questioning its responses.

The relational dimension is not decorative in a sociocultural account either. Wood et al. [24] count recruitment and frustration control among the tutor's functions, meaning that holding the learner in the task is part of scaffolding rather than a pleasant addition to it. Whether a chatbot can perform that function, and whether students want it to, is a question our data relates to directly.

3 Method

3.1 The course and the AI chatbot

As part of a bigger institutional project to investigate the use of AI both for staff and students, an AI chatbot was developed for an introductory programming course. To explore how an AI chatbot that had integrated course material

would work we chose an online course, Introduction to programming, that had a substantial amount of textual material. The course is a 7.5 ECTS Bachelor level course that makes use of JavaScript as its programming language. It is offered both online and campus-based with the same learning outcomes. The chatbot was fed the entire course material for the online course. It makes use of gpt-4o, and the user interface is almost identical to the interface of OpenAI's ChatGPT. The researchers configured Temperature, Top P, and System Message.

3.2 The pedagogical ground for the LLM teacher system message

Rooted in the sociocultural tradition, the LLM teacher was prompted to demonstrate dialogic teaching, as this style may enhance agency [16] and problem-solving skills [7]. Dialogic teaching is also known as "interthinking": the way more than one person (or person-LLM) reason and refine ideas collectively through dialogue [15]. Communication patterns cultivating problem-solving skills are associated with "exploratory talk", including questions such as "what do you think?" and "why do you think that?", and turns including maybes [7]. We also instructed the LLM teacher to scaffold students in learning JavaScript within their zone of proximal development [4], and to test students' level of knowledge.

3.3 Data collection

Data collection was done in 3 instances Spring and Autumn 2025. The first data collection was done in April, with 4 student assistants who had taken the course. They got to test the AI chatbot for 2 hours, prompting theoretical and practical questions, before the interviews started. One student was interviewed individually, the rest as a focus group. The second data collection was done in May and consisted of interviews with two students taking the course, who were given access to the AI chatbot three weeks before the interviews started. The course had ca. 20 registered students, and we recruited the respondents by sending a message to all enrolled studenten through their LMS.

The third data collection was done in November with three focus group interviews with students taking the course consisting of 2, 4 and 3 students respectively. The respondents were recruited by one of the researchers in a work session after a lecture. The course had ca. 500 registered students. The students were given access to the AI chatbot in September, and were given a programming task 3 days before the interviews. The task also included instructions on how to use the chatbot, and how to collaborate as a group.

All interviews used semi-structured interview guides, and lasted around 30-40 minutes. The data was collected using a speech-to-text tool and note-taking. For the first data collection the configuration of the AI chatbot was set with both Temperature and Top P at 0.10. For the second and third data collection the AI chatbot was set with Temperature at 0.5, and Top P at 0.8.

3.4 Thematic analysis

To analyze our data, we followed Braun and Clarke's thematic analysis [3]. This included familiarizing ourselves with the data, and generating codes and identifying themes in iterations. Since we wanted to explore how the students used and perceived the chatbot, we chose a bottom-up and semantic approach. One of the authors was mainly responsible for creating codes. Together with a second author the codes and themes for the first dataset were discussed, before moving on to coding and creating themes for the second and third dataset.

3.5 Ethics

We received approval from the Norwegian Agency for Shared Services in Education and Research (SIKT, reference number 799969 with latest update of approval 23.04.2025). Our respondents provided written informed consent, were anonymised, informed how the collected data would be used, and explained their right to withdraw without academic consequence.

4 Findings

The findings chapter presents the themes and codes from the three datasets, presented separately and descriptively, and analysed in the discussion section.

4.1 Data set 1: Themes for interviews with four student assistants

Pedagogical utility and learning support. Our respondents view AI as a meaningful resource for active learning, understanding and programming-related problem-solving, describe it as an integrated part of their studies, and report daily use in their courses. One participant described AI as a "sparring partner": "Learning to spar with it, asking for details and steps, where you can follow up with questions about the small steps... I believe that is a wonderful manner of learning." The chatbot was used to understand concepts, to debug, to improve code quality and to get feedback on code, and it greatly reduced the time needed to find answers.

AI vs. human support. Participants held that AI and human support serve different functions. AI is valued for immediate answers, while help from student assistants (SAs) could involve waiting, and teachers sometimes hold back the answer, don't go in-depth, or don't explain properly. Teachers and SAs, on the other hand, may motivate, adapt to students' needs, and guide problem solving instead of simply giving a solution, so respondents viewed AI as a complement rather than a replacement. The AI's personality was seen as neutral and academic, but also friendly because it was always available and one didn't need to fear critique back.

Technical performance and behavioral risks. Participants generally believed the chatbot answered within the course syllabus, but raised concerns

about incorrect and incomplete answers, repeated prompting, answer length, interface design and responsiveness in longer chats. The most prominent concern was dependency, laziness and shortcut-oriented learning: AI may be used only to obtain answers if the goal of its use isn't clear, and because it is always available, some students may postpone their learning. Fairness, legitimacy and institutional boundaries were also discussed. Suggested improvements included adaptive learning modes switching between direct answers and deeper learning, proactive behaviors such as asking questions and providing guided exercises, and features such as adaptive feedback, weakness identification and long-term learner modeling.

4.2 Data set 2: Themes for interviews with two students

Comparison to other chatbots. The two respondents had not used chatbots extensively before, but had tried Copilot, Google Search AI, and a previous chatbot provided by the school. The study's chatbot had the great advantage of keeping to the syllabus, where the others confused them with answers that were irrelevant, outside the syllabus or far beyond their level; they had sometimes given up on those, but not on the study's chatbot. One issue was that response time decreased as the chat got longer.

Learning with the AI chatbot. The students expressed that the chatbot worked well for learning programming and mostly used it to get in-depth understanding of code and concepts. They asked questions like "What does an object consist of?", and questions concerning specific techniques such as .value, this, parseInt, and return. They also used it for an overview of the topics, one comparing it to an encyclopedia. Some topics were covered well enough in the course material and therefore AI use was not needed. The respondents appreciated the chatbot's ability to give code with explanation, and help in debugging. One pasted code in and had it explained, asking follow-up questions about what the code was for, the difference between techniques, and how it could be improved, and sometimes asked the chatbot to integrate techniques they wanted to explore into existing code. Reflecting on pitfalls, they noted that learning requires doing things oneself, and that awareness of the goal of one's AI use is essential. One respondent said the chatbot is so efficient they stopped Googling: they saved time, but something was lost, since Googling through different sources one also learns surrounding topics.

Trust and personality of the chatbot. The students did not experience errors from the chatbot, and had no direct distrust of it, although one expressed that one shouldn't trust a chatbot blindly, just as one wouldn't a peer. They differed regarding personality. One thought the chatbot was friendly since it asked if they had more questions or wanted more explanation, and helped them see a direction when stuck. The other saw the personality as neutral, explaining things in the style of a lecturer good at presenting things step by step, and "straight to the point", just as they wanted.

4.3 Data set 3: Themes for interviews with three focus groups

Learning with the chatbot. The majority of our respondents reported that the chatbot too quickly gave a full solution, even when they just asked for an in-depth explanation. Prompting it to only guide on how to solve the task worked initially, but after a while it again gave full solutions. The one respondent who didn't face this issue had used the Remember function that makes the chatbot remember specific user prompts. Our students noted that it felt like cheating to get a full solution, and how hard it is not to use one once given. The feeling of mastery disappears as they do not find the solution themselves, and one said they sometimes felt dumb when realizing the solution was simple. They wanted the chatbot to ask whether they want the full solution or parts of it before responding.

Experiences with errors and code that didn't work differed. Some noted that instead of guiding them, the chatbot just corrected the code. Others noted how good it was at finding errors and at explaining and suggesting different manners of solving them, and some were simply grateful for the solution, as it saved them much time.

When the students managed to prompt the chatbot so that it didn't just give the solution it became of great value. The chatbot generated a recommended step-by-step plan, and a list of techniques to use explaining why and how. After the chatbot had first created the plan, the dialogue worked smoothly and they got incremental feedback as they proceeded. Getting a code skeleton to use as a fundament was also mentioned as important.

As a contrast, some students experienced that later in the process the explanations and code sometimes weren't enough to increase understanding. They acknowledged that the precision and level of detail of the prompt impacted the quality of the answer, and some mentioned that for the chatbot to stay focused they needed shorter code snippets and at most two sentences. One student had asked the chatbot in six different ways about something and still didn't understand it better. Another expressed that getting alternative manners of solving something sometimes made the subject feel harder, as opposed to getting the one right manner.

The students thought the chatbot worked well for learning theory, although one thought it stayed on the surface on some topics; another disagreed, having made it go deeper after prompting properly. It also worked well for explaining code they pasted in, and when students said they did not understand a code example, it could provide alternative versions they found easier. Some noted that it sometimes had difficulties providing the right code. Several also pointed out that answers were generally long, one calling 5-6 paragraphs very long; long code answers could be experienced as messy, and one had received a too long plan for solving a task.

Reflecting on how to best learn with the chatbot, the students said it is important to take time to reflect over the answers, ask "why-follow-up questions", and wait until you really need it. Good learning occurs when exposed to alternative explanations and manners of problem-solving, whereas using it just to get

quick answers is a pitfall. They noted that chatbots are so efficient they stopped reading literature, and worried about the chatbot "substituting their brain". One student's technique was to state in the prompt what they had learned and how they planned to use it in a task, to get feedback on their understanding. Another proposed that the chatbot should ask "What have you tried so far?" or "Show me what you have done!?"

Working as a group with the chatbot. The respondents collaborated as groups, discussing code to understand it, and one group read code line by line together. One student pointed out they relied more heavily on the chatbot due to the interviews, usually referring more to what the teacher had said. They got different answers and types of answers on the same prompt, but saw this as beneficial since it opened for discussions. They also solved the tasks in different manners with different techniques, and used this as an opportunity to explain their code to the others; learning by teaching. One group decided for one task to use a single prompt to get a step-by-step guide as the starting-point for all.

Interacting with the chatbot. Describing how they perceived the chatbot, they used words such as academic, professional, serious, formal, not personal, not funny, monotonous, and going straight to the point, and one compared it to a lecturer they have had. Seriousness was appreciated by some and expected of something offered by an educational institution, to make it credible. Some felt they were communicating with a very knowledgeable person that didn't want to get to know them. For some this worked, as they didn't want a relationship with a chatbot, but others expressed that had it behaved nicer they would seek it out more, as coziness was important.

Others perceived the chatbot as positive, helpful, polite, forever patient, never getting angry - comparable to a friendly person eager to help; one noted it said sorry when it did something wrong, and felt it proper to be polite back. One wanted encouraging comments like "You are beginning to get control over this!" alongside feedback on what to improve, adding that a cozy chatbot would not make them feel so dumb and would lighten the atmosphere, and one suggested that it should be possible to configure the personality. Comparing with ChatGPT, respondents explained that ChatGPT has more personality because it uses emojis and expressions such as "This is wonderful!", and even tries to be funny.

The students reflected that interaction with a human is more fluid and casual. One said "One doesn't have smalltalk with a chatbot". The chatbot was very technical, which is OK if one understands the technical terms, but sometimes respondents didn't know how to prompt it to get the answer they needed, and were aware they could get completely wrong answers if not sufficiently specific; a human adapts better. It was expressed that the chatbot ought to be able, like a human supervisor, to select important parts of code, point to them on the screen, and ask questions for reflection, and to check what the student knows with questions like "Have you used array and loop before?" One student therefore noted that if the chatbot had been a human it would have been a bad supervisor. Lastly, a good human supervisor gives students a spark to work and explore; the chatbot did not spark exploration and they had to take the initiative themselves.

Comparison to other chatbots. This study's chatbot was better than ChatGPT at relevancy and specificity, staying within the syllabus and saying when something was outside it, whereas with ChatGPT some students got techniques outside the syllabus, and even other programming languages. One noted that this study's chatbot more clearly said what was good or bad, compared to ChatGPT, which seemed positive to everything. Some respondents nevertheless thought ChatGPT, Copilot and Claude better for programming. One had tried ChatGPT's "study and learn" mode, noting it would not just give a solution. A couple mentioned that ChatGPT lets you choose answer styles and adapts as you use it; one explained "after knowing you more it gets more engaged and wants to play ball." A final advantage is uploading images and files, which for visual tasks becomes a manner of "pointing" at what is needed.

5 Discussion

This article reports on the exploratory phase in a pedagogical design project. We asked broadly: in what ways do students perceive the AI chatbot, customized for dialogic interaction, as supporting their learning process? Our discussion of the findings is divided into two sections: conflicting epistemic beliefs, and mastery, motivation and metaphorical prompting.

5.1 Conflicting epistemic beliefs

Students in our study seem to struggle with a recurring ambivalence toward AI-supported learning. They broadly agreed that the chatbot was useful for theoretical questions, code examples, structured explanations and distinctions between concepts, and that it became a sparring partner whose feedback improved code quality. At the same time, respondents across the data sets express fear of superficial learning and of a false sense of security in being knowledgeable, and describe the need to engage actively, to think before asking AI, and to ask follow-up questions. We use ambivalence here for this coexistence of positive and negative evaluations rather than cognitive dissonance in Festinger's sense [10], which would imply that the inconsistency itself produces discomfort and motivates attempts to reduce it.

Several previous studies report the same concerns about laziness, weakened thinking skills and declining code quality [25, 18, 12]. These perceptions are implicitly based on a cognitive view of learning [7] focused on the individual, and reflect the cognitive psychology of offloading, where learners outsource cognitive load to AI tools [23]. Judging AI's effect on learning solely from a cognitive-analytic stance will, however, hide other aspects of the learning situation. Another possibility is that the students align with a socially sanctioned expectation about appropriate AI use. From a Goffman-perspective, their responses may be understood as presentations conforming to what he calls the "officially accredited values" of society [11, p. 35]. That they tend to judge AI as a tool weakening their cognitive functioning could suggest they have adapted to public discussion

about cheating in higher education, or to values expressed by teachers in the classroom. Hence, students' feeling of learning or not may reflect a cultural phenomenon and a systemic problem [8], more than an accurate reflection of their cognitive development.

Offloading is discussed in contradictory terms: AI seems to enhance self-efficacy [26] and conceptual understanding [12], while there are serious concerns about students' uncritical dependence on it and offloading of cognitive effort [6]. A recent contribution is that we need to unpack the distinction between adaptive and maladaptive offloading rather than putting the two in conflicting categories. Wang and Zhang [23] suggest there may be two cognitive pathways in play simultaneously, vigilance and offloading, and that some offloading may be strategic and reallocate efforts for higher-order thinking. Hence, the fear of AI offloading found in the current study may counteract the students, as they do not explore the benefits. The ambivalence students express is also an emotional issue, raising the need for motivation for learning to happen, and conversely the feeling of mastery as the ground for motivation and psychological well-being [2].

5.2 Mastery, motivation and metaphorical prompting

Previous studies have shown that AI tutors in programming can help students master coding at a higher level, faster and more easily, which can lead to increased self-efficacy and motivation [25, 26]. Self-efficacy can be connected to Vygotsky's ideas of learning in the zone of proximal development [4], which we use in the system message in our AI tutor. The ZPD is the point where you can do something only with assistance from someone more skilled; that assisted mastery may lead to self-efficacy and agency to act [2], and an LLM tutor might serve as the assistant [4]. However, although our AI tutor was given a system message designed to support agency-focused dialogue, participants did not explicitly report any signs of such behavior or an increase in motivation, and said they would prefer a human as a motivator for exploration. We can only speculate about the reasons. One possibility is technical: the system message may not be specific enough to explain concepts such as ZPD and dialogic teaching, and the tutor's tendency to give extensive, informative answers with little emotional undertone indicates that. Another is students' understanding of motivation: if their conception connects to behaviorist views, such as the need for positive feedback and reward, this tutor did not provide it. It might nevertheless have supported motivation by scaffolding mastery, and the findings show that it did so on some occasions; the successful stories also seem to be driven more by student prompting than by the system instructions.

Some students in our study were concerned about the AI tutor giving them the solution too quickly, even when specifically asked not to. We do not know the correct answer to this problem, but a system-message pedagogy may erode over a prolonged conversation. Only one student escaped the problem, by storing the scaffolding instruction in the chatbot's "Remember" function so that it persisted across turns. It is a single case, but it suggests that the weakness lies in the persistence of the instruction rather than in the chatbot's capacity to

withhold an answer. A consequence of getting a solution is that students lose the opportunity to solve the task themselves and feel mastery. They felt they had cheated, but admitted it was hard not to use a solution once given. In ZPD terms, a delivered solution is not assistance [22]. However, this also raises questions about the students' prompting skills. When the teacher told them not to ask for the solution, they added a negative: do not give me the solution. The AI tutor needs more specific instructions on how to respond, both in the system instruction and in students' prompts.

The students who reported AI behavior comparable to a teacher's scaffolding were those in the group dataset, asked to collaborate on a task. They prompted the tutor for a "code skeleton" or a "step-by-step plan", called it a "sparring partner", and identified very good plans including techniques to use and explanations. Using the plan, they worked iteratively, receiving feedback at each step. What they describe is recognisable as scaffolding in the sense of Wood, Bruner and Ross [24]: the techniques the task required, with an explanation of why each was needed, followed by a plan the student could work through. In their terms this is a reduction of the degrees of freedom in the task combined with direction maintenance, and it is the behaviour the system message was written to produce. This finding suggests that students who succeed in prompting for scaffolding unintentionally use metaphors connected to scaffolding. If students learn to prompt by asking for a "code skeleton" or other metaphors that re-signify scaffolding, they may, in practical terms, gain the advantage of an AI tutor scaffolding them within their zone of proximal development.

In contrast to findings by Groothuijsen et al. [12], the group corpus also seems to value peer programming. Students working together noticed that when several students fed the tutor the same prompt, the AI's different answers let them discuss advantages, disadvantages, and alternative solutions. Through a sociocultural lens, the collective work of negotiation and talk is central to learning [15], suggesting that a sociocultural approach to AI peer-programming can be fruitful for future research.

6 Conclusion

This exploratory study points to a need for further research, for example into how epistemic beliefs influence students' own judgments of their learning outcomes, and into adaptive offloading, since students' beliefs here seem to be informed by the cognitive load concept in accounting for both negative and positive outcomes. Heavy cognitive load may not be the answer to good learning. We argue that research on AI-assisted learning needs to take the sociocultural aspects of learning more into account, and that the adaptive offloading concept in particular should be explored in that context, for example through the Zone of Proximal Development. We also touch on the cultural aspects of learning and judgments of cognitive functioning, and how public discourse on AI in higher education and teachers' beliefs may be reflected in students' views. It is central for teachers and LLM developers in education to collaborate to align the AI tool with learning

cultures and reduce students' dissonance from uncertainty, for instance by planning AI-supported peer programming and using metaphors to scaffold system instructions and AI programming didactics.

Our findings may suggest a number of things to someone building or deploying a course-specific chatbot in an introductory programming course. Binding the chatbot to the syllabus works: respondents in the second and third datasets preferred it to general-purpose chatbots because it stayed inside the course's language, level and techniques, and said so when something lay outside them. Pedagogy should not rest in the system message alone, since it depending on the setup and other things different may degrade over a prolonged conversation; withholding needs support at the architectural level, or at minimum a persistent instruction memory of the kind the one successful student used. Conversational style is better left as a student-controlled setting than decided at design time, as our respondents did not agree about warmth and we see no reason to expect another group to. Finally, the activity matters as much as the tool: the model's variability became productive when students worked in groups on a shared task with explicit instructions for collaboration.

Limitations. Our data are self-reports, collected shortly after use, by different student groups, and are the accounts of what students noticed. They are not evidence about learning outcomes.

References

1. Baig, M.I., Yadegaridehkordi, E.: Chatgpt in the higher education: A systematic literature review and research challenges. International Journal of Educational Research **127**, 102411 (2024)
2. Bandura, A.: Toward a psychology of human agency. Perspectives on Psychological Science **1**(2), 164–180 (2006)
3. Braun, V., Clarke, V.: Using thematic analysis in psychology. Qualitative Research in Psychology **3**, 77–101 (01 2006)
4. Cai, L., Msafiri, M.M., Kangwa, D.: Exploring the impact of integrating ai tools in higher education using the zone of proximal development. Education and Information Technologies **30**(6), 7191–7264 (2025)
5. Deng, H., Chen, H., Dong, Y.: Do ai chatbots improve students' learning performance in programming education? evidence from a meta-analysis. Journal of Educational Computing Research **64**(5) (2026)
6. Deng, R., Jiang, M., Yu, X., Lu, Y., Liu, S.: Does chatgpt enhance student learning? a systematic review and meta-analysis of experimental studies. Computers & Education **227**, 105224 (2025)
7. Dohn, N.B., Kafai, Y., Mørch, A., Ragni, M.: Survey: Artificial intelligence, computational thinking and learning. KI - Künstliche Intelligenz **36**(1), 5–16 (2022)
8. Engeström, Y., Sannino, A.: From mediated actions to heterogenous coalitions: four generations of activity-theoretical studies of work and learning. Mind, Culture, and Activity **28**(1), 4–23 (2021)
9. Ferguson, C., van den Broek, E.L., van Oostendorp, H.: Ai-induced guidance: Preserving the optimal zone of proximal development. Computers and Education: Artificial Intelligence **3**, 100089 (2022)

10. Festinger, L.: A Theory of Cognitive Dissonance. Stanford University Press, Stanford, CA (1957)
11. Goffman, E.: The Presentation of Self in Everyday Life. Doubleday, Garden City, NY (1959)
12. Groothuijsen, S., van den Beemt, A., Remmers, J.C., van Meeuwen, L.W.: Ai chatbots in programming education: Students' use in a scientific computing course and consequences for learning. Computers and Education: Artificial Intelligence **7**, 100290 (2024)
13. Labadze, L., Grigolia, M., Machaidze, L.: Role of ai chatbots in education: systematic literature review. International Journal of Educational Technology in Higher Education **20**(56) (2023)
14. Lan, Y., Liu, S., Chen, H.: The effects of responsiveness, perceived warmth, and anthropomorphism on university students' use of conversational ai for learning support: a chain mediation analysis based on s-o-r framework. Frontiers in Psychology **Volume 17 - 2026** (2026). https://doi.org/10.3389/fpsyg.2026.1769964
15. Littleton, K., Mercer, N.: Interthinking: Putting Talk to Work. Routledge, London (2013)
16. Omland, M., Rødnes, K.A.: Building agency through technology-aided dialogic teaching. Learning, Culture and Social Interaction **26**, 100406 (2020)
17. Richter, S., Kishore, S., Piven, I., Dodd, P., Bate, G.: Chatbots in tertiary education: Exploring the impact of warm and competent avatars on self-directed learning. British Journal of Educational Technology **56**(5), 2102–2124 (2025)
18. Silva, C.A.G.d., Ramos, F.N., de Moraes, R.V., Santos, E.L.d.: Chatgpt: Challenges and benefits in software programming for higher education. Sustainability **16**(3) (2024). https://doi.org/10.3390/su16031245, https://www.mdpi.com/2071-1050/16/3/1245
19. Sonlu, S., Bendiksen, B., Durupinar, F., Güdükbay, U.: The effects of embodiment and personality expression on learning in llm-based educational agents (2024)
20. Stojanov, A.: Learning with chatgpt 3.5 as a more knowledgeable other: an autoethnographic study. International Journal of Educational Technology in Higher Education **20**(35) (2023)
21. Tran, M., Balasooriya, C., Semmler, C., Rhee, J.: Generative artificial intelligence: the 'more knowledgeable other' in a social constructivist framework of medical education. npj Digital Medicine **8**(430) (2025)
22. Vygotsky, L.S.: Mind in Society: The Development of Higher Psychological Processes. Harvard University Press, Cambridge, MA (1978)
23. Wang, S., Zhang, H.: Pedagogical partnerships with generative ai in higher education: how dual cognitive pathways paradoxically enable transformative learning. International Journal of Educational Technology in Higher Education **23**(1), 11 (2026)
24. Wood, D., Bruner, J.S., Ross, G.: The role of tutoring in problem solving. Journal of Child Psychology and Psychiatry **17**(2), 89–100 (1976)
25. Yilmaz, R., Karaoglan Yilmaz, F.G.: Augmented intelligence in programming learning: Examining student views on the use of chatgpt for programming learning. Computers in Human Behavior: Artificial Humans **1**(2), 100005 (2023)
26. Yilmaz, R., Karaoglan Yilmaz, F.G.: The effect of generative artificial intelligence (ai)-based tool use on students' computational thinking skills, programming self-efficacy and motivation. Computers and Education: Artificial Intelligence **4**, 100147 (2023)