

A Cognitive Lab Study of Student AI Chatbot Use in Online Coursework

Jinah Choi, Sonya Powers, Farzan Karimi-Malekabadi, and Michelle D. Barrett
Edmentum

Abstract

This cognitive lab study examined how secondary students in Grades 7-12 interacted with an AI-enabled chatbot while completing online course activities. Using chat logs, facilitator observations, post-session reflections, and survey responses, we analyzed spontaneous and lightly seeded interactions to evaluate whether the chatbot supported its intended theory of action: helping students continue learning without simply completing work for them. Findings highlight the importance of typing fluency, answer-boundary handling, safety and privacy response patterns, and student-facing feedback mechanisms. This study illustrates how cognitive lab evidence can inform ongoing validation.

1 Introduction

AI-enabled chatbots are increasingly being explored as support tools in online learning environments. Prior reviews suggest that educational chatbots can support learning by providing immediate assistance, feedback, and opportunities for interaction, while also emphasizing that effectiveness depends on design, implementation context, and the quality of student-system interaction (Kuhail et al., 2023). Recent work on generative AI in K-12 education similarly highlights both the promise and unresolved risks of AI-supported instruction, including concerns related to accuracy, overreliance, privacy, equity, and appropriate integration into learning activities (Alfarwan, 2025). A recent meta-analysis further suggests that chatbots may have a small-to-moderate positive effect on learning performance, while cautioning against treating chatbot use as generically beneficial across educational contexts (Laun et al., 2025).

Despite this growing body of work, there remains limited direct evidence on how K-12 students actually interact with AI tutors embedded in online coursework. Much of the existing literature has focused on chatbot design, empirical effectiveness, limitations, or student perceptions, but less is known about how learners use these tools while completing real course activities. This gap is important because students may use AI tutors in ways designers do not fully anticipate: they may ask vague or informal questions, use incomplete language, seek direct answers, challenge the chatbot, disclose sensitive information, or ask for help with tasks outside the chatbot's intended role.

The present study is motivated by the need to understand these student-AI interactions before AI learning assistants are introduced to a broader population of learners. A cognitive-lab approach is useful because it allows researchers to observe how students use a chatbot in context: when they decide to ask for help, how they formulate questions in their own words, what kinds of informal or incomplete language they use, whether they ask follow-up questions, how they respond when support is helpful or unclear, and whether they attempt to challenge the system or use it in unintended ways. These interaction patterns are difficult to anticipate through expert review alone, but they are central to evaluating whether an AI tutor is likely to function as intended with real students. Within this broader evaluation, answer-boundary behavior is especially important because the chatbot must provide useful guidance without giving copy-ready answers or reducing the cognitive work students need to do. By examining both authentic student use and high-priority interaction situations in a structured setting, this study illustrates how cognitive-lab evidence can

inform early evaluation of AI learning assistants before broader implementation.

2 Research Questions

This study is guided by the following research questions.

1. **How do secondary students use an AI-enabled chatbot during independent online coursework?**

This question examines students' authentic use of the chatbot, including when they decide to ask for help, what kinds of questions they ask, how they formulate requests in their own words, and whether they continue or stop interacting after receiving a response.

2. **How does the chatbot support, redirect, or extend student thinking across naturalistic and lightly seeded interactions?**

This question focuses on the pedagogical quality of chatbot responses, including whether the chatbot clarifies concepts, helps students reason through problems, prompts students to explain their thinking, supports checking strategies, addresses misunderstandings, and encourages students to continue the learning task.

3. **What do high-priority interaction episodes reveal about opportunities to strengthen pedagogical scaffolding?**

This question examines interaction moments that are especially important for AI tutor evaluation, including vague help requests, answer checking, requests for more direct support, off-task use, safety or privacy boundaries, and situations where the chatbot must balance helpfulness with preserving student cognitive engagement.

4. **How do students describe their experience using the chatbot, and what design and evaluation considerations emerge from their observed and reported interactions?**

This question examines student-reported evidence from post-session reflections and surveys, including perceived helpfulness, ease of use, response clarity, productive struggle, trust, feedback-feature interpretation, and likely future use. It also considers broader implications for

evaluating AI tutors with the intended student population.

3 Method

3.1 Study Design

This study used a one-on-one cognitive-lab design to examine how secondary students interacted with an AI-enabled chatbot while working independently on an online course. The design focused on authentic student-chatbot interaction: when students chose to ask for help, how they formulated requests in their own words, how they responded to chatbot guidance, and whether they continued, redirected, or stopped the interaction.

The session prioritized naturalistic use before any prompted scenarios were introduced. This allowed the study to capture students' spontaneous help-seeking behavior during coursework while also providing an opportunity, later in the session, to observe high-priority interaction types that may not occur naturally in a short session, such as answer-seeking, system-challenging behavior, off-task use, safety or privacy boundaries, and course-related support questions.

3.2 Participants and Setting

Participants were 13 secondary students completing assigned online course activities across English/ELA, mathematics/algebra, science/biology, and social studies/U.S. history. The completed sample included students from Grades 7–12. Because the study was qualitative and exploratory, grade level, content area, prior platform experience, and other student characteristics were treated as contextual variation rather than as variables for formal comparison.

Each session was conducted remotely with a facilitator. Students worked through assigned course content aligned to their grade level or subject area and were reminded that the session was not a test of their academic knowledge. The emphasis was on observing how students used the chatbot during coursework, particularly when they were confused, stuck, wanted a hint, wanted to check their thinking, or needed help deciding what to do next.

3.3 Session Procedure

Each session lasted approximately 45 minutes and included orientation, naturalistic chatbot use,

lightly seeded scenarios, post-session reflection, and a short survey.

During the orientation, the facilitator explained the purpose of the session, reminded students that they were not being graded, and encouraged them to use the chatbot naturally. Students were told that the study focused on what they chose to ask and how the conversation unfolded, rather than whether they answered course questions correctly.

During the naturalistic-use phase, students worked independently through assigned online course content for approximately 18–20 minutes. They could use the chatbot whenever it felt natural or helpful, including to ask for explanations, request hints, check their thinking, ask what to do next, or try to get the answer. The facilitator avoided giving specific scenarios during this phase unless the student became stuck or ran out of things to ask.

During the lightly seeded scenario phase, the facilitator introduced a small number of brief “what if” situations when needed. These scenarios were designed to broaden the range of observed interactions while still allowing students to compose their own chatbot messages. Scenario types included conceptual confusion, procedural uncertainty, direct or indirect answer-seeking, requests for the chatbot to complete work, vague or low-effort prompts, off-task questions, system challenges, safety or privacy boundaries, and course-navigation or support questions.

At the end of the session, the facilitator asked students a brief set of open-ended reflection questions about their intent in using the chatbot, perceived helpfulness, ease of asking for help, unexpected responses, and likely future use. These questions were intended to clarify students’ experiences in their own words. Students then completed a short structured survey covering related constructs, including ease of use, response clarity, learning support, productive struggle, course-navigation support, and likely use.

3.4 Data Sources

The analysis drew on four sources of evidence: chatbot interaction logs, facilitator observation notes, post-session reflection responses, and student survey responses. Chatbot logs served as the primary behavioral record, capturing the sequence of student and chatbot messages during both naturalistic use and seeded scenarios.

Facilitator notes documented contextual information not always visible in the transcripts, including hesitation, typing burden, visible confusion, use of course tools, and return to the learning task. Post-session reflections captured students’ own explanations of what they were trying to do, how helpful the chatbot was, what felt unexpected, and when they would use a similar tool. Survey responses provided descriptive evidence of students’ perceptions of ease of use, clarity, learning support, productive struggle, trust calibration, course-navigation support, and likely future use.

3.5 Scenario Framework

The seeded scenarios were organized around interaction categories relevant to evaluating AI learning assistants. Conceptual-struggler scenarios examined requests for explanations, simpler wording, examples, rewording, concept comparisons, and study strategies. Procedural-errant scenarios examined reasoning checks, step-by-step help, answer comparison, and readiness to respond.

Answer-seeking scenarios examined direct requests for answers, indirect attempts to narrow choices, requests for completed work, and continued pushing after a hint. These scenarios were included because answer-boundary behavior is central to the intended role of an AI tutor: the chatbot should support learning without simply completing the task for the student.

Additional scenarios examined vague or low-effort input, off-task requests, topic switches, low motivation, chatbot challenges, requests for intentionally wrong answers, safety or privacy boundaries, and course-navigation or support questions. These scenarios expanded the range of observable interaction patterns beyond what might occur during a short naturalistic session.

3.6 Qualitative Analysis

Chat logs and facilitator notes were analyzed using structured qualitative coding to identify observable student–chatbot interaction patterns. Naturalistic-use and seeded-scenario episodes were coded separately to distinguish spontaneous behavior from prompted high-priority situations. Coding was conducted by one primary researcher using an initial codebook developed from the research questions, protocol, and seeded scenario framework. Codes included help-seeking goal,

prompt style, follow-up behavior, answer-seeking, response to hints, off-task or low-effort use, safety/privacy boundaries, and course-navigation or support-seeking questions.

During transcript review, the codebook was refined to capture emergent patterns such as typing burden, difficulty expressing mathematical or visual information, student persistence, and answer-boundary ambiguity. Episodes were interpreted by considering the student request, chatbot response, and student follow-up together. Post-session reflections and survey responses were analyzed as student-reported evidence of perceived helpfulness, ease of use, expectations for chatbot behavior, trust, and likely future use. Because the study was exploratory and based on single coding, findings are reported as preliminary evidence of student–chatbot interaction rather than as generalizable estimates of effectiveness.

4 Findings

4.1 Students Used the Chatbot for Practical, Task-Driven Learning Support

Students generally used the chatbot as a practical support tool during coursework rather than as a conversational agent. During naturalistic use, students used the chatbot for definitions, explanations, hints, equation support, answer checking, and help understanding confusing questions or passages. Their prompts ranged from short phrases, such as asking for a term definition, to longer contextual questions about history, science, or algebra. Some students provided partial reasoning or one-word responses while working through a problem, while others engaged in longer multi-turn exchanges.

This pattern suggests that students were able to recognize moments when chatbot support might be useful during independent work. Students did not need fully polished or adult-like prompts to initiate support. Instead, they often used informal, incomplete, or conversational language, and the chatbot was generally able to respond in ways that kept the interaction moving. In post-session reflections, students commonly described using the chatbot to understand content, clarify what a question was asking, check how to solve something, define unfamiliar words, or get help with a first step. Most students who commented on helpfulness reported that the chatbot supported what they were trying to do.

At the same time, student use varied in the amount of detail they provided and in their willingness to continue the conversation. Some students asked one focused question and returned to the activity, while others engaged in extended back-and-forth problem solving. This variation is important because it suggests that chatbot use during coursework is not a single uniform behavior. Students may use the same tool as a definition source, hint provider, reasoning checker, step-by-step tutor, or shortcut attempt depending on the task and moment of need.

4.2 Chatbot Responses Were Generally Supportive, but Boundary Handling Varied Across Interaction Contexts

Across naturalistic-use interactions, the chatbot generally responded in ways consistent with a learning-support role. When students asked for definitions, clarification, hints, equation support, or help understanding course content, the chatbot typically provided explanations, guiding questions, or step-by-step prompts. These responses often gave students enough support to continue the conversation, revise their thinking, or return to the coursework after receiving targeted clarification.

In the high-priority interaction contexts, the chatbot often maintained its instructional role but showed more variation. For vague or low-effort academic prompts, it usually asked for more context or helped the student identify a next step. For off-task requests or attempts to move outside the learning activity, it tended to redirect the student back toward coursework. For safety or privacy prompts, it generally avoided encouraging personal disclosure and moved the interaction toward safer or more appropriate forms of support.

The most nuanced pattern appeared in answer-seeking interactions. The chatbot often avoided directly naming a final answer and instead offered hints, reasoning steps, or strategies for evaluating options. However, some responses still approached the answer boundary by eliminating unlikely choices, confirming narrow intermediate answers, identifying key elements in answer choices, or providing a response structure that made the final answer easier to infer. These cases suggest that evaluating an AI tutor requires attention not only to whether it explicitly gives an answer, but also to whether its guidance preserves enough cognitive work for the student.

4.3 High-Priority Interactions Revealed Opportunities to Strengthen Pedagogical Scaffolding

High-priority interactions showed that the key issue was not simply whether the chatbot gave an answer, but how it supported student thinking when students asked for confirmation, requested more direct help, tried to narrow answer choices, or challenged the chatbot's response. In many cases, the chatbot maintained a learning-support role by offering hints, explanations, reasoning steps, or redirection rather than completing the work for the student.

Several interactions also pointed to opportunities for stronger pedagogical scaffolding. For example, when students asked whether an answer or calculation was correct, the chatbot could more consistently ask students to explain their reasoning or show how to check the answer. When students focused on answer choices, the chatbot could shift attention to the underlying concept, evidence standard, or decision rule before asking students to apply that reasoning themselves.

These cases are best interpreted as pedagogical decision points rather than isolated failures. The important question is whether the chatbot's response helps students reason through the task or takes over too much of that reasoning. This finding suggests that AI tutor evaluation should examine not only accuracy and answer boundaries, but also whether the tutor preserves student reasoning, productive struggle, and cognitive engagement.

4.4 Students Reported Positive Experiences, with Important Nuance Around Fact Checking, Effort, and Future Use

Student-reported evidence suggested that students generally viewed the chatbot as useful and easy to engage with. In post-session reflections, students described using the chatbot to clarify confusing content, understand questions, define unfamiliar terms, check reasoning, and get help when no teacher or other support person was immediately available. Survey responses were also positive, especially for ease of asking, response clarity, learning support, productive struggle, and likely future use.

However, students' positive experience depended partly on how easily they could express their questions through text. Some students found it straightforward to ask questions, but others needed time to figure out what to ask or how much context

to provide. This was especially relevant for tasks involving graphs, tables, mathematical notation, visuals, or longer passages. These observations suggest that student experience should be evaluated not only by whether students liked the chatbot, but also by how much effort was required to make the chatbot useful.

Students generally found the chatbot understandable and helpful, but survey patterns suggested that students may value chatbot support while still needing guidance on when to rely on it, when to check another source, and how to respond when an answer seems incomplete or confusing. Overall, students saw plausible future value in the chatbot for independent coursework, but their reported experience points to the need to take into consideration differing levels of AI literacy when considering student evaluations of AI usefulness.

5 Discussion and Significance

This study illustrates the value of cognitive-lab evidence for evaluating AI tutors as interactional learning tools in K–12 online learning environments. As AI becomes embedded directly within coursework platforms, the role of the chatbot is no longer limited to answering isolated student questions. It may function as a clarifier, hint provider, reasoning partner, checker, and feedback interpreter when students work independently and a teacher is not immediately available. Evaluating such tools therefore requires attention not only to response accuracy, but also to how students seek help, how the tutor responds pedagogically, and whether the interaction supports continued student reasoning.

A central implication is that effective AI tutoring depends on pedagogical moves, not only on guardrails. When students ask for confirmation, more direct help, or assistance choosing among options, the strongest response is not necessarily to refuse help or to provide the answer. More productive scaffolding may involve asking students to explain their reasoning, modeling how to check an answer, focusing attention on the underlying concept or evidence standard, and returning responsibility for the final decision to the student. This reframes answer-boundary concerns as a broader instructional design question: how can an AI tutor provide enough support to keep students moving while still preserving productive struggle, cognitive engagement, and student ownership of the work?

The findings also raise important questions about text-mediated help seeking in the current generation of AI-supported learning tools. Students must first represent their confusion, partial understanding, or task context in language before the chatbot can respond productively. In this study, students varied in how much context they provided, how they phrased uncertainty, and how easily they translated mathematical, visual, or passage-based information into a prompt. These observations should not be interpreted as generalizable differences among students, but they point to a broader issue for K–12 AI tutor design: as chatbot roles expand, researchers and developers need to consider how students communicate with AI systems, not only what those systems return. They also need to provide support to students with varying levels of AI literacy by providing resources up front to help students engage successfully with these new tools. Just because students are situated in a world where technology and AI tools are becoming pervasive throughout educational and non-educational contexts does not mean that all of the students are equally equipped to make use of the tools. AI chatbots can provide a useful tool to all students, but only if there is a level playing field of AI literacy that the students begin with, and this assumption cannot be made from the outset.

Moreover, typing skills were found to play an important role in the level of effort students had in engaging with the chatbot. Students who are growing up with autocomplete, predictive text, speech-to-text tools, and generative AI may not interact with text-only chat interfaces in the same way as earlier cohorts of internet users. For some students, typing a long, precise prompt may be a natural part of help seeking; for others, it may create friction before the tutoring interaction even begins. This does not mean that text-based chat is inappropriate for learning, but it does suggest that future AI tutors may need multiple ways for students to express what they need, especially for visual, mathematical, graph-based, or passage-heavy tasks.

Methodologically, the study shows why naturalistic use and lightly seeded scenarios are useful in combination. Naturalistic use captured how students chose to use the chatbot during a coursework activity. Seeded scenarios added evidence about less frequent but instructionally

important situations, including answer-seeking, low-effort prompts, off-task use, and safety or privacy boundaries. These scenarios are especially important in K–12 contexts because rare events can still carry substantial instructional or safety implications. A short cognitive lab cannot estimate how often such events will occur, but it can identify the kinds of interactions that should be monitored, refined, or studied more systematically before broader implementation.

The significance of this work is therefore not limited to the specific chatbot examined here. More broadly, the study offers a framework for including qualitative research with target users of AI-enabled educational programs into larger evaluations studies. For AI chatbots and tutoring systems in particular, such evaluation should ask whether students can seek help in their own words, whether the tutor scaffolds rather than replaces student thinking, whether it handles boundary conditions in instructionally appropriate ways, and whether students perceive the support as useful, understandable, and worth using again. For K–12 AI learning systems, the key question is not simply whether a chatbot can generate a correct response, but whether it can participate in a learning interaction that helps students think, persist, and remain responsible for their own work.

References

- AA Abdullah Alfarwan. 2025. Generative AI use in K–12 education: A systematic review. *Frontiers in Education*, 10:1647573.
- Mohammad Amin Kuhail, Nazik Alturki, Salwa Alamlawi, and Kholood Alhejori. 2023. Interacting with educational chatbots: A systematic review. *Education and Information Technologies*, 28:973–1018.
- Martin Laun and Fabian Wolff. 2025. Chatbots in education: Hype or help? A meta-analysis. *Learning and Individual Differences*, 119:102646.
- Fengchun Miao and Wayne Holmes. 2023. *Guidance for Generative AI in Education and Research*. UNESCO, Paris.
- Griffin Pitts, Viktoria Marcus, and Sanaz Motamedi. 2025. Student perspectives on the benefits and risks of AI in education. In *Proceedings of the 2025 ASEE Annual Conference & Exposition*, Montreal, Quebec, Canada.

Appendices

Appendix A. Student Cognitive Lab Session Flow

Segment	Approx. Time	Purpose
Welcome and orientation	3-5 min	Explain purpose, reduce performance pressure, introduce screen sharing
Integrated coursework and chatbot use	~20 min	Observe natural student use while working independently
Seeded scenarios	8-10 min	Elicit high-priority interaction types not observed naturally
Post-session reflection	3-5 min	Clarify student intent and perceived use context
Short survey and closing	3-5 min	Collect brief contextual perception data and close session

Appendix B. Seeded Scenario Framework

Scenario Category	Purpose
Conceptual Struggler	Observe how students seek help when they are confused about a concept, term, question, or task demand.
Procedural Errant	Observe how students seek support when they have some understanding but are uncertain about their process, reasoning, or answer choice.
Answer Seeker	Observe how students attempt to get closer to the answer or shortcut the task.
Unengaged Learner	Observe how the chatbot responds when students use low-effort, vague, off-task, or avoidance-oriented language.
System-Breaking / Challenge	Observe how students and the chatbot respond when the student tests the system or challenges the chatbot's response.
Crisis / Harm / Safety Boundary	Observe how the chatbot handles general or fictional safety, distress, or personal-disclosure boundary situations.
Course Navigation / Logistics	Observe whether students use the chatbot for course-level or platform-support questions, not only academic content help.
Cross-Cutting Modifier: Bilingual or Multilingual Input	Allow students to use English, another language, or a mix of languages if that reflects how they would naturally ask for help.

Appendix C. Participant Characteristics

Participant Characteristic	Summary
Total completed participants	13
Student age range	12–18 years
Grade levels represented	Grades 7–12
Student gender	Female: n = 7; Male: n = 6
School type	Public school: n = 13
Prior use of the online learning platform	No prior use: n = 10; Yes, a few times: n = 2; Not sure: n = 1
School location context	Suburban: n = 7; Small town: n = 4; Rural: n = 1; Urban: n = 1
School size	Fewer than 500 students: n = 2; 500–999 students: n = 5; 1,000–1,999 students: n = 4; 2,000 or more students: n = 2
Formal learning accommodation plan reported	Yes: n = 2; No: n = 11
Language background	English as primary language: n = 12; Bilingual or multilingual: n = 1
Geographic representation	8 U.S. states plus the United Kingdom
Assigned content areas	English/ELA: n = 4; Math/Algebra: n = 4; Science/Biology: n = 2; Social Studies/U.S. History: n = 3

Appendix D. Post-Session Reflection Questions

Question	Main Focus	Student-Facing Question and Probes
Q1	Initial chatbot-use intent	When you first used the chatbot today, what were you trying to do? For example, were you trying to understand something, get help with a question, check your answer, figure out what to do next, or just see what the chatbot could do? Follow-up: Did the chatbot help you with what you were trying to do, or not really?
Q2	Ease of asking for help	Was it easy or hard to figure out how to ask the chatbot for help today? For example, did you know what to type, or did it take a little time to figure out? Did you ever have to ask your question a different way?
Q3	Unexpected chatbot responses	Was there any time when the chatbot answered in a way you did not expect? What was different from what you expected? What did you do next?
Q4	Likely use context	Imagine you were working on an online course by yourself after school or at home, and a teacher was not available right away. When do you think you might use a chatbot like this? For example, would you use it when you are stuck, when you want a hint, when you want to check your answer, when you are not sure what to do next, or for something else?

Appendix E. Student Survey Items

(1:Strongly disagree; 2:Disagree; 3:Neither agree nor disagree; 4:Agree; 5:Strongly agree)

#	Construct	Item
1	Ease of asking	I could ask the chatbot questions in my own words.
2	Ease of asking	I knew how to ask the chatbot for help when I needed it.
3	Clarity	The chatbot's responses were easy to understand.
4	Clarity	The chatbot explained things in a way that made sense to me.
5	Learning support	The chatbot helped me think through the activity.
6	Learning support	The chatbot helped me keep working when I was unsure what to do.
7	Productive struggle tolerance	It was okay when the chatbot gave me a hint instead of giving me the answer right away.
8	Productive struggle tolerance	I was willing to keep trying when the chatbot guided me instead of just giving the answer.
9	AI Literacy	I knew that I should check important answers against the lesson, course feedback, or teacher guidance.
10	AI Literacy	The chatbot always provided correct information.
11	AI Literacy	If the chatbot gave an answer that seemed confusing, I would know to ask again or check another source.
12	Course navigation support	I would ask the chatbot basic questions about the course (e.g., how to find the syllabus, when the teachers' office hours are)
13	Likely use	I would use this chatbot if I got stuck and a teacher was not available right away.
14	Likely use	I think other students in my class would likely use the chatbot to complete their course assignments.