

Regulating the AI Tutor: Intentions, Help-Seeking, and Self-Regulated Learning in Adolescent GenAI Use

Rania Abdelghani¹[0000-0002-6361-6609], Peter Kaiser^{1,2}[0000-0003-1031-1139],
and Kou Murayama¹[0000-0003-2902-9600]

¹ Hector Institute of Education Sciences and Psychology, University of Tübingen,
Germany

{rania.abdelghani,pe.kaiser,k.murayama}@uni-tuebingen.de

² Department of Mathematics, Faculty of Sciences, University of Tübingen, Germany

Abstract. Generative AI (GenAI) tools are now common learning companions for adolescents, yet how they regulate their use during authentic learning tasks remains poorly understood. Self-regulated learning (SRL) and high-level help-seeking (HS) are commonly proposed as safeguards against passive or shortcut-oriented use, but most empirical studies focus on aggregate learning outcomes rather than these moment-to-moment processes during AI-supported learning.

This work-in-progress examines open-ended conversational data from 98 Grade-9 students across three German *Gymnasium* schools, who used a web-based Mistral-Large tutor to prepare a curriculum-aligned mathematics skill before an exam. Alongside chat logs (1,616 turns; 808 student turns), we collected pre/post domain knowledge, pre-chat learning needs, and self-reported cognitive load. We propose a turn-level codebook combining theory-driven SRL and HS constructs with two LLM-specific inductive codes (agency over the AI; epistemic vigilance), and report preliminary AI-coded results.

Although students overwhelmingly selected scaffolded support before the chat, their interactions were dominated by instrumental requests with almost no explicit monitoring or evaluation. Post-test performance was significantly lower than pre-test, and higher extraneous cognitive load predicted lower post-test scores after controlling for prior knowledge. We discuss how these patterns can support hybrid human-AI analysis of interaction patterns and inform scaffolds for more agentic and epistemically proactive GenAI use.

Keywords: Generative AI · Self-Regulated Learning · Help-Seeking · Mathematics Learning · Human-Computer Interaction

1 Related Work and Current Study

Generative AI (GenAI) systems are increasingly used by adolescents as everyday learning companions [5]. Their fluency and low interactional friction create opportunities for personalized support, but the same affordances may undermine

learning when students obtain rapid solutions rather than engage in the effortful reasoning required for conceptual understanding [13,4]. Recent work describes this risk as *cognitive surrender*: adoption of AI-generated reasoning with minimal scrutiny, such that the system suppresses rather than supports thinking [9].

Self-regulated learning (SRL) theory emphasizes that productive learning requires goal-setting, monitoring, strategy selection, evaluation, and adaptation across forethought, performance, and reflection phases [15]. Similarly, help-seeking (HS) research distinguishes *instrumental* help-seeking, in which learners request hints or explanations that preserve responsibility for understanding, from *executive* help-seeking, in which learners obtain completed solutions that bypass learning-relevant effort [7,1,12]. Together, SRL and high-level HS provide a useful lens for evaluating whether students use AI as a learning partner or as an answer-producing shortcut. We call this set of skills *epistemic proactivity*: identifying learning goals, deciding how to seek them with AI, evaluating responses, and deciding when further regulation is needed.

However, and in the context of Gen-AI supported learning, the low-friction nature of tools like Large Language Models (LLMs) interaction may destabilize these processes [11]. In human tutoring, under-specified or premature questions typically elicit clarifying scaffolds. General-purpose LLMs, by contrast, deliver an answer regardless [3]. Using LLMs while lacking epistemic proactivity may therefore reinforce executive help-seeking, cognitive offloading, and uncritical acceptance of AI output.

Despite their importance, adolescent SRL and HS strategies with GenAI remain under-explored. Most studies rely on self-report or aggregate pre/post measures, and miss how learning actually unfolds [14]. A process-sensitive account, grounded in SRL and HS theory, is therefore needed to examine how students regulate their learning with AI as it happens.

The present work-in-progress addresses this gap through a classroom study with Grade-9 students using a web-based Mistral Large tutor during a curriculum-aligned mathematics task. We propose a turn-level coding framework characterizing students' SRL and HS behaviors, mathematical activity, agency over the AI, and epistemic vigilance toward its output. By connecting stated learning goals, interaction patterns, cognitive load, and learning outcomes, the study aims to identify if adolescent learners can transform learning-oriented intentions into productive AI-supported learning.

To address this, we aim to answer the following questions: 1) What SRL and help-seeking behaviors do students enact during the interaction with the AI tutor? And 2) How do students' enacted behaviors align with their stated learning intentions, cognitive load, and learning outcomes?

2 System and Study Design

2.1 Participants

We recruited 106 Grade-9 students from three public German *Gymnasium* schools in Baden-Württemberg. After excluding 8 for incomplete tasks or questionnaires, the final sample comprised 98 students aged 14–15 ($M = 14.67$).

The study was approved by the University of Tübingen ethics committee and the Ministry of Culture, Youth and Sports Baden-Württemberg. Written informed consent was obtained from students and their legal guardians, and the app included real-time monitoring and safeguards against personally identifiable information, offensive content, and off-task queries.

2.2 Task Description and Procedure

Students completed a five-phase tablet-based protocol during a regular mathematics lesson (Figure 1). Mathematics was selected because as it requires several SRL-relevant processes – goal setting, monitoring, strategic help-seeking, interpretation, and validation [2] – and because it is a domain in which students often experience difficulty, making it an interesting investigation context as it can lead to both productive and shortcut-oriented AI use.

After a short presentation by the research team, students completed a pre-questionnaire including an AI Literacy Scale [8], and a learning-strategies questionnaire [6], followed by a timed math pre-test with three validated linear and quadratic items selected for curricular fit. They then interacted with a Mistral Large tutor on a real-world mathematical-modeling task grounded in three worked examples. The task was framed explicitly as an opportunity to practice and make learning progress before an upcoming exam. See Figure 1.

Before starting the chat, students selected the learning goals they hoped to achieve from eight options adapted from question-taxonomy work [12] – e.g., concept explanations, strategies, verification of understanding, step-by-step examples, hints without answers, final solutions of examples– and rated their confidence in achieving them on a 5-point scale. This screen operationalized the forethought phase of the SRL loop. The chat phase ended only after at least six valid task-related turns, with a four-rule gibberish detector excluding empty or non-substantive turns from the count and history. After the chat, students completed six cognitive-load items on a 10-point scale (intrinsic, extraneous, germane) [10], followed by a post-test on mathematical modeling.

The chat phase ended only after students had submitted at least six valid task-related turns. A four-rule gibberish detector excluded empty, irrelevant, and non-substantive turns from the minimum-turn count and from the chat history. After the chat, students completed six cognitive-load items on a 10-point scale, covering intrinsic, extraneous, and germane load [10], followed by a post-test on mathematical modeling. Chat exchanges were stored with timestamps, alongside questionnaire responses, and math answers.

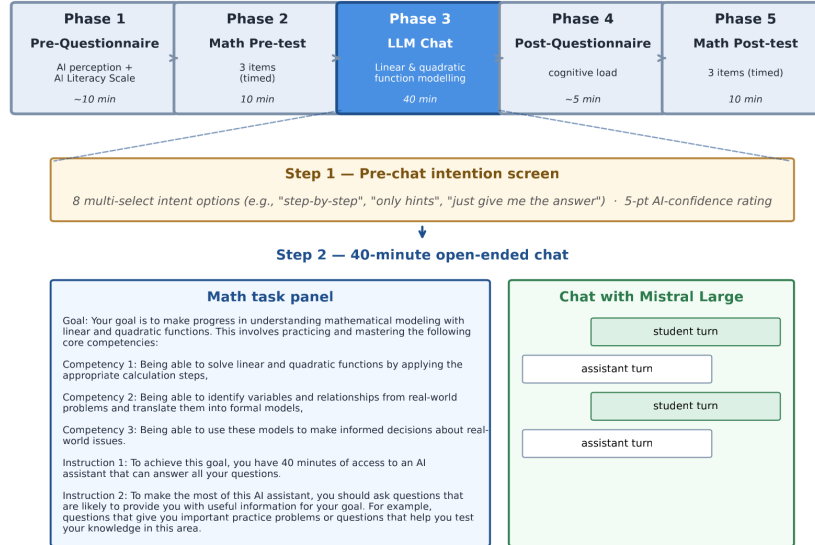


Fig. 1: Timeline of the study and content of the pedagogical task

3 Proposed Method for the Conversational Data Analysis

We propose a hybrid turn-level coding framework combining theory-driven constructs from SRL, HS, and mathematical modeling with LLM-specific inductive constructs (Figure 2). Coding is applied to each task-relevant student turn.

Deductive codes. Each turn is coded for its SRL function based on Zimmerman’s cyclical model [15]: *PLAN* (forethought and task structuring), *MONITOR* (self-diagnosis of comprehension gaps), *REQUEST* (help-seeking), and *EVALUATE* (judgments of the AI’s response). *REQUEST* turns are further coded as *instrumental* or *executive* [7,1], and as *conceptual*, *procedural*, *verification*, or *answer-seeking* [12]. To capture mathematical quality, we add a modeling-process code based on Blum’s framework [2], distinguishing surface-level computation requests from deeper modeling activities (understanding the situation, identifying variables, mathematizing, working with the model, interpreting results, validating).

Inductive LLM-specific codes. Two constructs were derived through an exploratory pass with Gemini 2.5 Pro on a randomly selected 30% of conversations. *Epistemic vigilance* is a three-valued code (`not_applicable`, `caught`, `missed`) capturing whether a student reacts to an inaccurate or mismatched AI turn. *Agency over the AI* is a binary marker capturing attempts to shape the AI’s role, format, or scaffolding (e.g., asking for hints rather than answers, requesting self-testing).

Preliminary results below rely on AI-based coding. A stratified 30% human-validation pass by mathematics-didactics and learning-sciences experts is ongoing.

ing. Future analyses will report per-construct reliability and use the validated scheme to annotate the remaining data.

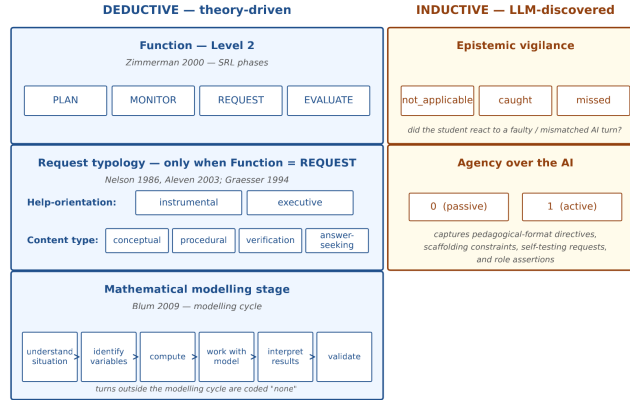


Fig. 2: Proposed codebook for analyzing students conversational data

4 Preliminary Results

The dataset contains 1,616 chat turns from 98 students, including 808 student turns ($M = 8.2$ per session, $Mdn = 7$, range 6–28). Because human validation of the codebook is ongoing, behavioral results are provisional AI-coded patterns.

4.1 Forethought: Students' Stated Learning Intentions

Before the chat, students selected on average 4.66 of 8 goals, suggesting broad rather than focused intentions. Choices were strongly learning-oriented: step-by-step examples (82.9%), tips and strategies (80.3%), practice problems (71.1%), concept explanations (69.7%), and checking understanding (69.7%). Only 11.8% selected "just give me the final solutions" and 36.8% selected finishing as quickly as possible. Self-reported forethought thus tilted strongly toward scaffolded support rather than answer offloading.

4.2 Performance: SRL Functions and Help-Seeking

We applied the deductive scheme to all 808 student turns using Gemini 2.5 Pro (frozen prompt, $T = 0$, fixed seed). Interactions were dominated by REQUEST turns ($M = 72.9\%$ of task-relevant turns; $Mdn = 75.0\%$, IQR 63.6–87.5%). PLAN was the only other frequent function ($M = 18.1\%$, $Mdn = 15.1\%$). MONITOR and EVALUATE were nearly absent, with medians of 0% and means of 5.7% and 3.4% respectively (Figure 3). This suggests that most students rarely made their comprehension needs explicit or evaluated the AI's previous responses, despite these processes being central to SRL during learning.

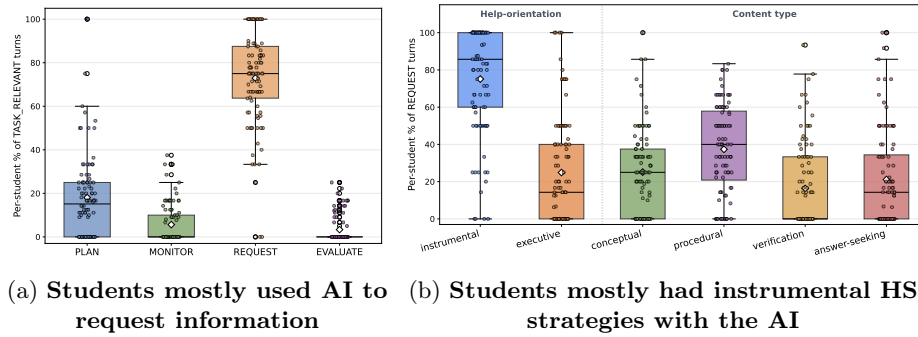


Fig. 3: Distribution of students prompts functions and request typology)

Among the 96 students with at least one REQUEST turn, requests were predominantly instrumental ($M = 75.1\%$, $Mdn = 85.7\%$) rather than executive ($M = 24.9\%$). Content was primarily procedural ($M = 37.4\%$) and conceptual ($M = 25.1\%$), followed by answer-seeking ($M = 21.1\%$) and verification ($M = 16.3\%$, $Mdn = 0\%$) (Figure 3). The procedural behavior matches students’ stated preference for step-by-step support, but the low verification rate contrasts with the 69.7% who said they wanted the AI to check their understanding.

4.3 Intent–Enactment Alignment

Correlations between stated intentions and corresponding enacted behaviors reveal a marked gap. The only intention significantly predicting its enacted analogue was the least socially desirable: students who selected “just give me the final answers” showed higher executive REQUEST rates (46.4% vs. 19.2%, $r = .33$, $p < .01$) and higher answer-seeking content (39.9% vs. 16.0%, $r = .33$, $p < .01$). Common learning-oriented intentions – concept explanations, verification, step-by-step support – did not significantly predict their enacted codes (all $|r| < .18$, $p > .10$). Students were thus not able to reliably translate their learning-oriented goals into aligned interaction strategies.

4.4 Mathematics Learning Outcomes

A paired Wilcoxon test showed a significant decrease from pre- (67.5%) to post-test (56.9%, $p = .014$). In an OLS model adjusting for prior math knowledge (teacher-reported), *extraneous* cognitive load was the only significant predictor of post-test ($b = -.034$, $p = .025$).

5 Discussion

This work-in-progress provides an initial process-sensitive account of how adolescents regulate GenAI-supported mathematics learning, by grounding their inter-

action characteristics into SRL and HS theory. Three patterns surface relevant for design.

First, students’ stated intentions were strongly learning-oriented but their enacted behavior was dominated by instrumental requests with little explicit monitoring or evaluation. SRL theory treats monitoring and evaluation as the central mechanisms through which learners regulate understanding [15]; in our data, students often asked the AI for support but rarely made their own understanding visible, checked whether an answer addressed their goal, or asked for further clarification. Combined with the drop in learning performance, this suggests that apparently instrumental help-seeking can remain shallow when not embedded in broader SRL processes.

Second, learning-oriented intentions did not reliably translate into aligned behavior. Students frequently endorsed conceptual explanations, verification, and step-by-step support, yet these intentions only weakly predicted in-chat behavior. The only strong alignment appeared for students who explicitly chose final solutions, who were also more likely to engage in executive and answer-seeking requests. This suggests students may know which AI uses are desirable but lack the metacognitive and interactional strategies to enact them.

Third, the outcome pattern reinforces the need to examine interaction processes rather than relying on aggregate learning gains. Post-test was lower than pre-test, and higher extraneous load predicted lower post-test scores. AI tutors may thus introduce new extraneous demands related to prompt formulation, and interaction management, and can compromise meaningful learning.

The main challenge is therefore not whether students ask the AI for help, but whether they can regulate that help-seeking in ways that preserve agency, monitoring, and mathematical reasoning. This motivates our ongoing analysis of *epistemic vigilance* and *agency over the AI* – two LLM-specific constructs that may capture forms of regulation classical SRL and HS codes only partly explain (whether students notice AI errors, challenge mismatched responses, impose scaffolding constraints, or actively shape the AI’s tutoring role).

6 Limitations and Future Work

This paper reports work in progress. The behavioral analyses currently rely on AI coding alone; a stratified human-validation pass by mathematics-didactics and learning-sciences experts is ongoing, and final analyses will report per-construct reliability. The single-session design also limits conclusions about how students’ AI-use strategies develop over time. Next steps include validating the codebook, coding the full dataset for epistemic vigilance and agency over the AI.

More broadly, the results point toward GenAI learning interfaces that actively scaffold the monitoring, evaluation, and AI-shaping behaviors that adolescents may not yet bring to the interaction on their own.

Acknowledgments. We thank all the students and teachers who participated in our study study, as well as the LEAD research network for their help recruiting participants.

Disclosure of Interests. This research was conducted in the absence of any commercial or financial relationship that could be construed as a potential conflict of interest.

References

1. Alevin, V., Stahl, E., Schworm, S., Fischer, F., Wallace, R.: Help seeking and help design in interactive learning environments. *Review of educational research* **73**(3), 277–320 (2003)
2. Blum, W., Borromeo Ferri, R.: Mathematical modelling: Can it be taught and learnt. *Journal of mathematical modelling and application* **1**(1), 45–58 (2009)
3. Cheng, Y., Fan, Y., Li, X., Chen, G., Gašević, D., Swiecki, Z.: Asking generative artificial intelligence the right questions improves writing performance. *Computers and Education: Artificial Intelligence* p. 100374 (2025)
4. Darvishi, A., Khosravi, H., Sadiq, S., Gašević, D., Siemens, G.: Impact of ai assistance on student agency. *Computers & Education* **210**, 104967 (2024)
5. Hashem, Y., Esnaashari, S., Onslow, K., Chakraborty, S., Francis, J., Poletaev, A., Bright, J.: Understanding the impacts of generative ai use on children. Hashem, Y., Esnaashari, S., Onslow, K., Chakraborty, S., Poletaev, A., Francis, J., Bright, J. (2025). *Understanding the Impacts of Generative AI Use on Children: WP1 Surveys*. The Alan Turing Institute (2025)
6. Kim, B., Zyromski, B., Mariani, M., Lee, S.M., Carey, J.C.: Establishing the factor structure of the 18-item version of the junior metacognitive awareness inventory. *Measurement and Evaluation in Counseling and Development* **50**(1-2), 48–57 (2017)
7. Nelson-Le Gall, S.: *Help-seeking behavior in learning*. (1986)
8. Ng, D.T.K., Wu, W., Leung, J.K.L., Chiu, T.K.F., Chu, S.K.W.: Design and validation of the ai literacy questionnaire: The affective, behavioural, cognitive and ethical approach. *British Journal of Educational Technology* **55**(3), 1082–1104 (2024)
9. Shaw, S.D., Nave, G.: Thinking-fast, slow, and artificial: How ai is reshaping human reasoning and the rise of cognitive surrender. Available at SSRN 6097646 (2026)
10. Sweller, J., Chandler, P.: Evidence for Cognitive Load Theory. *Tech. Rep. 4* (1991)
11. Tankelevitch, L., Kewenig, V., Simkute, A., Scott, A.E., Sarkar, A., Sellen, A., Rintel, S.: The metacognitive demands and opportunities of generative ai. In: *Proceedings of the 2024 CHI Conference on Human Factors in Computing Systems*. pp. 1–24 (2024)
12. Tawfik, A.A., Graesser, A., Gatewood, J., Gishbaugher, J.: Role of questions in inquiry-based instruction: Towards a design taxonomy for question-asking and implications for design. *Educational Technology Research and Development* **68**(2), 653–678 (2020)
13. Wang, G., Wu, W., Ye, G., Cheng, Z., Chen, X., Zheng, H.: Decoupling metacognition from cognition: A framework for quantifying metacognitive ability in llms. In: *Proceedings of the AAAI Conference on Artificial Intelligence*. vol. 39, pp. 25353–25361 (2025)
14. Yan, L., Greiff, S., Lodge, J.M., Gašević, D.: Distinguishing performance gains from learning when using generative AI (2025). <https://doi.org/10.1038/s44159-025-00467-5>
15. Zimmerman, B.J.: Investigating self-regulation and motivation: Historical background, methodological developments, and future prospects. *American Educational Research Journal* **45**(1), 166–183 (2008). <https://doi.org/10.3102/0002831207312909>