The Effect of AI Hint Transparency on User Retention in an Educational Platform

Maria Zinchenko, Konstantin Bagrationi and Olga Tunkevichus
Graduate School of Business, HSE University, Frankfurt am Main, Germany

mzinchenko@hse.ru
kbagrationi@hse.ru
oagordienko@hse.ru

Abstract: Generative artificial intelligence (AI) tools are increasingly embedded in coding courses, but their educational value depends on how this assistance is delivered to the student (Brown et al., 2025). This study tests whether revealing that clues are generated by AI - indicating their origin, naming their limitations, and requiring verification - affects the early retention of teenagers in an online course on coding. Developed a dedicated web platform providing a two-week self-paced Python introductory course, and conducted a supervised field experiment involving 93 high school and college students which were divided into three transparency conditions by means of a fixed alternating sequence: clues without mentioning that they are generated by AI (A), indicated clues with the verification instruction (B) and indicated clues with the verification instruction and the mandatory microphlexy (C) Four behavioral metrics were also calculated: week-2 retention, module-1 completion rate, stuck rate, and hint-to-progress time. The more transparent formats were associated with significantly higher retention and completion rates, as well as lower trapping frequency and faster progress after the hint, with the most complete C format having the strongest effect. The results suggest that the tooltip design, which combines source disclosure, information about AI failure, test queries and microflexy, is associated with lower early drop-off without changing the content of the course. This can be seen as a component of the measurable design mechanism.

Keywords: Artificial Intelligence in Education, AI Transparency, Learner Retention, Programming Education, Learning Analytics, Instructional Design

1. Introduction

Generative AI models are rapidly being integrated into educational scenarios, including school and additional programming courses. For teenage students, this environment is particularly sensitive, as learning to code requires constant feedback. AI cues simultaneously speed up tasks and increase the risk of uncritical misdecisions (Pitts et al., 2025). On the other hand, normative and methodological recommendations emphasize human-centric design, clear rules for using AI, and transparency of the role of an instrument. Moreover, the formats for defining AI content are deemed methodologically unreliable (Miao & Holmes, 2023; Lodge, 2024).

Against this background, there is a scientific challenge: determining which formats of AI tooltips help reduce dropout and increase retention in teen programming courses. Although the effects of AI assistants in programming have already been studied, the results are mixed. Some studies show faster completion of tasks, but they also point to risk: newcomers may take erroneous cues without analysis. This may weaken their understanding of how solutions work (Akçapınar & Sidan, 2024). The results tend to vary depending on the type of task, discipline and prior training, which reinforces the need for careful consideration of AI usage rules.

In this article, transparency is considered as a controlled pedagogical mechanism that shapes the behavior of the student. We define three operational formats of transparency, implement them as interface conditions within the platform and evaluate their relationship with retention metrics. The result of this work is a proven, operationalized transparency of AI-tips and a set of design recommendations for educational course developers.

2. Literature Review

Generative AI in school education is described in literature as a rapidly expanding but methodologically heterogeneous field for which clear regulatory frameworks are particularly important (Lin & Tan, 2025). The authors of the articles note a lack of field and long-term studies at the level of primary and further education, as well as insufficient knowledge of the mechanisms through which AI influences educational behavior (Yusuf et al., 2024). The recurring concern is that when key reasoning and validation steps are delegated to AI, learning activity becomes less autonomous: fewer attempts, less error analysis, and less reflection (Alfarwan, 2025).

In the context of programming, empirical work reports that chat bots are used primarily to understand and receive implementation guidance, i.e., as a tool for operational support (Groothuijsen et al., 2024). At the same time, quasi-experimental data on high school students show that those who used an AI assistant showed lower

performance and self-efficacy compared to the control group (Yang et al., 2025). Process-oriented systems such as CodeHelp (Liffiton et al., 2023) and CodeAid (Kazemitabaar et al., 2024) show that usage regulation can be designed at the interface level by offering explanations, guidance, and tips without providing ready-made solutions. This is perceived as beneficial by both students and teachers. The educational effect of an assistant depends on how it is embedded in the pedagogical design of the course (Humble et al., 2024; Garcia, 2025).

Transparency is understood here as revealing the role of AI in educational activities, and the conditions under which it is used. What matters most is how the explanation is structured in the interface and whether it corresponds to users' cognitive expectations (Naiseh et al., 2021; Balasubramaniam et al., 2023; Xu et al., 2025). The simple label "generated by AI" may not be sufficient if it is not accompanied by limitations and clear verification instructions (European Commission, 2022; Perkins et al., 2024).

Retention in educational literature means continuing to engage the student in key stages of the program (Shaikh & Asif, 2022). Surveys show that retention depends on a combination of student characteristics, course structure, and digital environment. Moreover, self-regulation, self-efficacy, motivation, and a sense of belonging play a prominent role (Buizza et al., 2024; Cefa & Zawacki-Richter, 2025). Therefore, the article considers the change in pedagogical design as a factor affecting student retention.

Finally, the adaptive hint posits support as a flexible intervention that should give hint without a ready answer (Doroudi et al., 2019; Maniktala et al., 2023). The wheel-spinning phenomenon, which is associated with a high risk of early exit, makes it important to recognize such patterns at an early stage (Shen et al., 2021). Despite many numerical studies, the literature lacks large-scale field experiments in school and extracurricular programming on the transparency of AI use. Moreover, few practical guidelines are offered for integrating transparency into the coaching platform interface. This gap is what motivates the present study.

3. Methodology

The empirical part of the work was organized as a controlled field experiment in an online self-study environment. The goal was to test how different formats of AI-lead transparency affect problem-solving and student retention. The independent variable was the format of the hint presentation; dependent variables were behavioral retention metrics.

Three transparency conditions were compared, differing in the way they were presented, and the content of the tasks was the same (Table 1). In condition A, students saw only the text of the hint, with no indication that it was generated by AI. In condition B, the hint was a short block stating that the aid had been generated by AI, that the model could be wrong, and that the solution should be tested by running tests with an added boundary case. In condition C, the same flagged hint was accompanied by a mandatory microphraseology, in which the student briefly recorded what he had changed in his decision.

Table 1: Definition of the transparency formats

Format	Interface	Research Meaning
A – basic untagged hint	Hint text only, with no marker that it is AI-generated and no verification instruction.	Control condition: help is present, but its origin and limitations are hidden.
B – marked up text and verification	A short block before the hint: (1) source – the hint is from AI; (2) limitation – AI may be wrong; (3) verification – a prompt to run tests with an edge case.	Tests the effect of labelling the source and providing verification instruction without additional load.
C – marked up text, verification reflection	The same block as in B, followed by a mandatory micro-action: filling in a short “what I changed” field.	Tests on whether minimal reflection strengthens the transparency effect through deliberate verification.

The hints were not presented as ready-made solutions. They were designed as a controlled training aid.

Four metrics were used to measure the results (Table 2). The main metric was week-2 retention – the student’s presence of at least one active event during the second week of the course. Module-1 completion rate recorded whether they had completed all 20 tasks of the first module. Stuck rate measured the proportion of tasks that showed crawling, such as repeated test runs without progress or repetition of the same error. The task was classified as a "crawler" when a student made at least three consecutive test submissions without improving the

number of tests taken within a 20-minute window. Hint-to-progress time was the interval between showing the hint and the first improvement in the results of the automated tests, excluding pauses of more than two hours. All metrics were first calculated at the individual student level and then aggregated to the group level, so that each participant entered a group value exactly once.

Table 2: Study metrics

Metric	Definition	Aggregation
Week-2 retention	1 – if the learner has at least one active event in week 2, else 0. Primary retention metric.	Share of learners with value 1 per group.
Module-1 completion rate	1 – if the learner completed all 20 tasks of module 1, else 0.	Share of learners with value 1 per group.
Stuck rate	Share of started tasks flagged as wheel-spinning, repeated unsuccessful attempts.	Mean share of tasks per group.
Hint-to-progress time	Time from hint display to the first improvement in test results in minutes.	Mean time per group.

The course was a two-week introduction to Python in a self-guided mode, divided into two consecutive week modules of 20 tasks each. The second module opened automatically seven days after registration regardless of progress, so that returning to course could be interpreted as a behavioral solution rather than as a technical consequence of completing all tasks. Each task combined metadata, short wording with built-in theory, code editor, automated tests including boundary cases, and a layer of adaptive hints. Hints were generated via the API of the large language model with a limited spread that prevented the issuance of full code. The platform was built on Python backend and react front end with a built-in code editor, with logging of events: task openings, test runs, queries and show prompts, microflexy, and successful completions.

The study involved 93 students aged 13-20 from high schools and colleges in one country of Eastern Europe, on a voluntary basis. When registering, the platform allocated each member to condition A, B or C using a fixed alternating sequence (A, then B, then C, with repetition), without manual intervention in the allocation. This choice was dictated by the platform's technical limitations at the time of launch. No check for group comparability regarding demographic characteristics and prior programming experience was conducted, as such data were not collected during registration. This limitation is discussed in section 6. The fact that the course was part of a scientific inquiry was revealed in the welcome dialogue, but differences between the conditions were deliberately not reported. The experiment lasted two working weeks - a horizon chosen because dropout in online courses is concentrated in the first weeks of recording (Yang et al., 2013).

Two hypotheses have been formulated. H1: More transparent formats of AI prompts will be associated with higher retention: conditions B and C expect higher week-2 retention and module 1 completion rate what in condition A. H2: In conditions B and C, students will be less likely to tow on tasks and faster to show first progress after receiving the hint than in condition A, as a more transparent format will affect the decrease of stuck rate and hint-to-progress time. For binary retention metrics, the chi-square Pearson criterion was used for fractions; for continuous and temporal metrics, the non-parametric Kraskel-Wallis criterion was used, as these variables are continuous.

4. Results

All four metrics were automatically calculated from logs, ensuring comparability between groups (Table 3).

Table 3: Metrics by experimental group

Metric	Group A	Group B	Group C
Participants	31	31	31
Week-2 retention	29.0%	45.2%	64.5%
Module-1 completion rate	25.8%	58.1%	67.7%
Stuck rate	19.5%	11.5%	9.7%
Hint-to-progress time (avg)	17 min	10 min	9 min

Week-2 retention increased from 29.0% in group A to 45.2% in group B and 64.5% in group C. Module-1 completion followed a similar pattern, rising from 25.8% in A to 58.1% in B and 67.7% in C - a gap of about 42 percentage points between the end conditions. These values confirm H1: the formats that reveal the origin of AI prompts and the need to validate them resulted in a higher week-2 retention and module-1 completion rate than the condition where the source and restrictions of the prompts remained hidden.

Process metrics confirm the mechanism through which transparency influences. Stuck rate decreased from 19.5% in group A to 11.5% in B and 9.7% in C, so that students in group C were about half as likely to repeat unsuccessful attempts, what students of group A. Hint-to-progress time fell from about 17 minutes in A to about 10 minutes in B and 9 minutes in C. The decrease in both metrics under more transparent conditions indicates that, That source labeling, together with the verification instruction, forms a more informed strategy for interacting with AI: instead of mechanically applying the hint, the student checks the proposed solution, which simultaneously reduces the likelihood of repeated errors and accelerates progress. These results confirm H2.

Statistical significance was assessed using the Pearson's chi-quadrate criterion for binary metrics and the Kraskel-Wallis criterion for continuous metrics (Table 4). All four metrics showed statistically significant differences between groups at $p < 0.05$.

Table 4: Statistical tests (A, B, C)

Metric	Group A	Group B	Group C	Test	Statistic	df	η^2	p-value
Week-2 retention	29.0% [16 -47%]	45.2% [29-62%]	64.5% [47-79%]	χ^2	7.8726	2	-	0.0195
Module-1 completion rate	25.8% [14-43%]	58.1% [41-74%]	67.7% [50-81%]	χ^2	11.9584	2	-	0.0025
Stuck rate	19.5%	11.5%	9.7%	Kruskal–Wallis H	8.9724	-	0.0775	0.0113
Hint-to-progress time	17 min	11 min	9 min	Kruskal–Wallis H	50.0291	-	0.5787	<0.001

The most pronounced effect was recorded for the hint-to-progress time metric ($H = 50.0291$, $p < 0.001$), indicating a significant influence of the hint format on the speed of subsequent problem solving. In addition to testing significance, effect sizes were calculated: for the Kruskal-Wallis test for the hint-to-progress time metric, η^2 was 0.5787, indicating a large effect. This also means that the differences in the time between receiving a hint and progress are explained by the experimental condition. For stuck rate, the effect size was 0.0775, a moderate but robust effect (Cohen, 1988; Tomczak & Tomczak, 2014). For week-2 retention and module-1 completion rate metrics, 95% confidence intervals were additionally calculated using Wilson's method (Table 4). The intervals of group A do not overlap with the intervals of groups B and C for any of the indicators, which confirms the stability of the differences regardless of the point estimate p-value.

Two other patterns are apparent. The increase in module 1 completion rate is greater between groups A and B, while the increase in week-2 retention is greater between groups B and C, suggesting that source marking supports retention at the first stage of the course, while reflection is more important for returning to the longer horizon. Moreover, group C, despite the additional step of reflection, shows the shortest time from hint to progress, which contradicts the common assumption that mandatory extra steps slow learning.

5. Discussion

The monotone increase in week-2 retention and module 1 completion rate from group A to group C supports the view that AI lead transparency functions as a pedagogical mechanism. This aligns with survey-based evidence that AI transparency is one of the factors students associate with trust and reduced concern in educational AI use (Al-Zahrani, 2024), although that study did not measure behavioral learning outcomes directly. The present study builds on this position by showing that simply labelling a source in format B already yields significant retention gains compared to control, whereas mandatory microlending in format C yielded numerically the highest results.

The reduction in stuck rate observed here is larger than in comparable randomized studies of AI feedback and counterproductive persistence. In a randomized controlled experiment with GPT-4 feedback on compiler errors, observed a reduction in wheel spinning from approximately 8.6% to 6.1%, it is a relative reduction of about 29%

($\tau = -0.134$) (Zhou et al., 2025). In the present study, the stuck rate drops from 19.5% in group A to 9.7% in group C, a relative reduction of about 50%. This difference may be explained by the difference in intervention design: in this article AI feedback was added on top of regular compiler messages without indicating its source or reliability, whereas in the transparency condition tested here, the source was clearly labeled, its limitations were stated, and active verification was required in the C format.

Because participants were recruited in three cities with unequal group sizes, part of the observed effect could, in principle, be due to the composition of the cohort by city or school, rather than the prompt format itself. Other participant characteristics, such as age, gender, and previous programming experience, were not collected, so the comparability of the groups on these parameters cannot be assessed post hoc. This does not invalidate the main result, but it limits confidence that the observed differences are due solely to the prompt format and not the sample composition.

The decrease in stuck rate in groups B and C can be read as a more conscious interaction with the hint. Similar findings are found in work on AI computers for programming, which guide the student's thinking without producing ready answers and thus support autonomy (Bassner et al., 2024). In the present study, a similar effect is achieved solely by the way the hint is designed, which expands the pedagogical design tools available to course developers.

The result hint to progress time was the most statistically significant: group C, whose members were supposed to fill in the reflection field, showed the shortest mean time before the first progress. This result is explained by metacognitive theories of learning programming, as reflexive queries after coding tasks correlate with improved student outcomes (Choi et al., 2023). Moreover, reflexivity helps students move from a passive take on the hint to active cognitive processing (Sahin & Noroozi, 2025). Micro-phraseology in condition C refers to precisely this: instead of applying the hint mechanically, the pupil perceives it, which accelerates progress.

The study conceptually treats transparency as three design solutions with different content, which allows us to identify individual elements of transparency. Empirically, it focuses on the adolescent audience, for which data are still scarce against the predominance of university samples (Meyer et al., 2024). Methodologically, it shifts from the detection of AI use to observed behavior based on retention metrics measured through logs in accordance with current regulatory guidelines for AI assistants (Miao & Holmes, 2023; European Union, 2024).

Practically, this distinction matters for course developers working with limited interface development budgets: most of the completion-rate gain is already achieved by format B, which is simple source labeling paired with a verification instruction, while the mandatory micro-reflection step in format C requires additional development and adds a step to the user journey, without increasing task completion time. For platforms where the cost of additional interaction steps is high, format B may be a reasonable minimal intervention, while format C serves as a reinforcement for courses that prioritize longer-term engagement over first-module completion alone.

6. Conclusion and Further Recommendations

In this study, a working digital platform was designed that realizes three formats of AI-transparency as interface conditions with full event reporting. This platform conducted a two-week quasi-experimental field study with alternating distribution involving 93 adolescent students. Across all four metrics - week-2 retention, module-1 completion rate, stuck rate, and hint-to-progress time - statistically significant differences were observed between groups, with the most complete C format, including microreflection, having the strongest effect. Students in this group were less stuck, moved faster after the trip and returned more often in the second week. Format B, with source marking and inspection instructions, primarily improved the completion of the first module. Thus, two components of transparency touch upon different facets of retention.

On this basis, a set of design recommendations is proposed:

- The AI clues must carry a clearly separated source token;
- The tooltip block must include a neutral statement that the model is not unerring;
- He must give specific instructions on the verification (run tests, add a boundary case);
- Where the most powerful effect is required, it should require minimal mandatory microleseplexia after the hint.

In this study, the delivery of AI clues was associated with higher early engagement in the course and less trailing than unmarked clues. Because the intervention combined several design elements, the results should not be interpreted as a mere delineation of transparency. A number of limitations outline these conclusions. The

sample was restricted in size and limited to adolescents in a specific geography, and the study measured early rather than long-term retention. Retention is also sensitive to external factors: student accessibility, motivation and irregular participation, as well as third-party support in the context of home-based independent learning, which cannot be ruled out entirely. Hints were generated automatically without expert validation of their pedagogical quality, and logs record behavior rather than depth of understanding. Furthermore, due to the lack of randomization and data on participants' baseline characteristics at the time of registration, it cannot be ruled out that the differences between groups partly reflect the composition of the cohorts rather than solely the effect of the prompt format. Promising areas for further research include broader age and geographic samples, longer observation horizons, mechanism testing in other disciplines, monitoring background variables through short surveys, Expert validation of the quality of tips and combination of behavioral logs with qualitative methods such as interviews or test tests.

Ethical Declarations

Ethical clearance was not required for the research

AI Declarations

AI (ChatGPT 5.5 Thinking) was used to enhance language quality

References

Akçapınar, G., & Sidan, E. (2024). AI chatbots in programming education: Guiding success or encouraging plagiarism. Discover Artificial Intelligence, 4(1), 87.

Al-Zahrani, A. M. (2024). Unveiling the shadows: Beyond the hype of AI in education. Heliyon, 10(9), e30696.

Alfarwan, A. (2025). Generative AI use in K–12 education: A systematic review. Frontiers in Education, 10, 1647573.

Bassner, P., Frankford, E., & Krusche, S. (2024). Iris: An AI-driven virtual tutor for computer science education. In Proceedings of the 2024 Conference on Innovation and Technology in Computer Science Education V. 1 (ITiCSE 2024) (pp. 394–400). Association for Computing Machinery.

Balasubramaniam, N., Kauppinen, M., Rannisto, A., Hiekkanen, K., & Kujala, S. (2023). Transparency and explainability of AI systems: From ethical guidelines to requirements. Information and Software Technology, 159, 107197.

Brown, N. C. C., Weill-Tessier, P., Leinonen, J., Denny, P., & Kölling, M. (2025). Howzat? Appealing to expert judgement for evaluating human and AI next-step hints for novice programmers. *ACM Transactions on Computing Education*. https://doi.org/10.1145/3737885

Buizza, C., Bertelli, G., Ghilardi, S., & Pioli, R. (2024). The role of self-efficacy, motivation, and connectedness in predicting dropout in online higher education. Education Sciences, 14(1), 67.

Choi, H., Jovanovic, J., Poquet, O., Brooks, C., Joksimović, S., & Williams, J. J. (2023). The benefit of reflection prompts for encouraging learning with hints in an online programming course. The Internet and Higher Education, 58, 100903.

Cohen, J. (1988). Statistical power analysis for the behavioral sciences (2nd ed.). Lawrence Erlbaum Associates.

Doroudi, S., Aleven, V., & Brunskill, E. (2019). Where's the reward? A review of reinforcement learning for instructional sequencing. International Journal of Artificial Intelligence in Education, 29(4), 568–620.

European Commission, Directorate-General for Education, Youth, Sport and Culture. (2022). Ethical guidelines on the use of artificial intelligence (AI) and data in teaching and learning for educators. Publications Office of the European Union.

European Union. (2024). Regulation (EU) 2024/1689 of the European Parliament and of the Council laying down harmonised rules on artificial intelligence (Artificial Intelligence Act). https://artificialintelligenceact.eu/

Garcia, M. B. (2025). Teaching and learning computer programming using ChatGPT: A rapid review of literature amid the rise of generative AI technologies. Education and Information Technologies, 30(12), 16721–16745.

Groothuijsen, S., van den Beemt, A., Remmers, J. C., & van Meeuwen, L. W. (2024). AI chatbots in programming education: Students' use in a scientific computing course and consequences for learning. Computers and Education: Artificial Intelligence, 7, 100290.

Humble, N., Boustedt, J., Holmgren, H., Milutinovic, G., Seipel, S., & Östberg, A. S. (2024). Cheaters or AI-enhanced learners: Consequences of ChatGPT for programming education. Electronic Journal of e-Learning, 22(2), 16–29.

Kazemitabaar, M., Ye, R., Wang, X., Henley, A. Z., Denny, P., Craig, M., & Grossman, T. (2024). CodeAid: Evaluating a classroom deployment of an LLM-based programming assistant that balances student and educator needs. In Proceedings of the CHI Conference on Human Factors in Computing Systems (CHI '24), Article 650. ACM.

Liffiton, M., Sheese, B., Savelka, J., & Denny, P. (2023). CodeHelp: Using large language models with guardrails for scalable support in programming classes. In Proceedings of the 23rd Koli Calling International Conference on Computing Education Research, Article 8. ACM.

Lin, X., & Tan, H. (2025). A systematic review of generative AI in K–12: Mapping goals, activities, roles, and outcomes via the 3P model. Systems, 13(10), 840.

Lodge, J. M. (2024). The evolving risk to academic integrity posed by generative artificial intelligence: Options for immediate action. Tertiary Education Quality and Standards Agency (TEQSA).

Maniktala, M., Chi, M., & Barnes, T. (2023). Enhancing a student productivity model for adaptive problem-solving assistance. User Modeling and User-Adapted Interaction, 33(1), 159–188.

Meyer, J., Jansen, T., Schiller, R., Liebenow, L. W., Steinbach, M. H., Horbach, A., & Fleckenstein, J. (2024). Using LLMs to bring evidence-based feedback into the classroom: AI-generated feedback increases secondary students' text revision, motivation, and positive emotions. Computers & Education: Artificial Intelligence, 6, 100199.

Miao, F., & Holmes, W. (2023). *Guidance for generative AI in education and research*. UNESCO. https://doi.org/10.54675/EWZM9535

Naiseh, M., Al-Thani, D., Jiang, N., & Ali, R. (2021). Explainable recommendation: When design meets trust calibration. *World Wide Web*, *24*(5), 1857–1884.

Perkins, M., Furze, L., Roe, J., & MacVaugh, J. (2024). The Artificial Intelligence Assessment Scale (AIAS): A framework for ethical integration of generative AI in educational assessment. Journal of University Teaching & Learning Practice, 21(2).

Pitts, G., Rani, N., Mildort, W., & Cook, E.-M. (2025). *Students' reliance on AI in higher education: Identifying contributing factors*. arXiv. https://arxiv.org/abs/2506.13845

Sahin, I., & Noroozi, O. (Eds.). (2025). Transforming education through generative artificial intelligence (GenAI): Innovations, challenges, and opportunities. ISTES.

Shaikh, U. U., & Asif, Z. (2022). Persistence and dropout in higher online education: Review and categorization of factors. Frontiers in Psychology, 13, 902070.

Shen, S., Liu, Q., Chen, E., Wu, H., Huang, Z., Zhao, Y., & Wang, M. (2024). A survey of knowledge tracing: Models, variants, and applications. IEEE Transactions on Learning Technologies.

Tomczak, M., & Tomczak, E. (2014). The need to report effect size estimates revisited: An overview of some recommended measures of effect size. Trends in Sport Sciences, 1(21), 19–25.

Xu, G., Murthy, S. V., & Jia, B. (2025). Enhancing intuitive decision-making and reliance through human–AI collaboration: A review. Informatics, 12(4), 135.

Yang, D., Sinha, T., Adamson, D., & Rosé, C. P. (2013). Turn on, tune in, drop out: Anticipating student dropouts in massive open online courses. In Proceedings of the 2013 NIPS Data-Driven Education Workshop.

Yang, T. C., Hsu, Y. C., & Wu, J. Y. (2025). The effectiveness of ChatGPT in assisting high school students in programming learning: Evidence from a quasi-experimental research. Interactive Learning Environments, 33(6), 3726–3743.

Yusuf, A., Pervin, N., Román-González, M., & Noor, N. M. (2024). Generative AI in education and research: A systematic mapping review. Review of Education, 12(2), e3489.

Zhou, Y., Pankiewicz, M., Paquette, L., & Baker, R. S. (2025). Impact of LLM feedback on learner persistence in programming. In B. Jiang et al. (Eds.), Proceedings of the 33rd International Conference on Computers in Education. Asia-Pacific Society for Computers in Education.