



## OPEN ACCESS

## EDITED BY

Eugène Loos,  
Utrecht University, Netherlands

## REVIEWED BY

Noble Lo,  
Lancaster University, United Kingdom  
Deepika Dhamija,  
Manipal University Jaipur, India

## \*CORRESPONDENCE

Oleg Shvets  
✉ oleg.shvets@taltech.ee

RECEIVED 07 May 2025

ACCEPTED 28 July 2025

PUBLISHED 08 August 2025

## CITATION

Shvets O, Murtazin K, Piho G and  
Meeter M (2025) Experiment with ChatGPT:  
methodology of first simulation.  
*Front. Educ.* 10:1624516.  
doi: 10.3389/feduc.2025.1624516

## COPYRIGHT

© 2025 Shvets, Murtazin, Piho and Meeter.  
This is an open-access article distributed  
under the terms of the [Creative Commons  
Attribution License \(CC BY\)](#). The use,  
distribution or reproduction in other forums is  
permitted, provided the original author(s) and  
the copyright owner(s) are credited and that  
the original publication in this journal is cited,  
in accordance with accepted academic  
practice. No use, distribution or reproduction  
is permitted which does not comply with  
these terms.

# Experiment with ChatGPT: methodology of first simulation

Oleg Shvets<sup>1\*</sup>, Kristina Murtazin<sup>1</sup>, Gunnar Piho<sup>1</sup> and  
Martijn Meeter<sup>1,2</sup>

<sup>1</sup>Department of Software Science, Tallinn University of Technology (TalTech), Tallinn, Estonia, <sup>2</sup>LEARN!  
Research Institute, Vrije Universiteit Amsterdam, Amsterdam, Netherlands

Providing timely and effective feedback is a crucial element of the educational process, directly impacting student engagement, comprehension, and academic achievement. However, even within small groups, delivering personalized feedback presents a significant challenge for educators, especially when opportunities for individual interaction are limited. As a result, there is growing interest in the use of AI-based feedback systems as a potential solution to this problem. This study examines the impact of AI-generated feedback, specifically from ChatGPT 3.5, compared to traditional feedback provided by a supervisor. The aim of the research is to assess students' perceptions of both types of feedback, their satisfaction levels, and the effectiveness of each in supporting academic progress. As part of our broader research agenda, we also aim to evaluate the relevance of the domain model currently under development for supporting automated feedback. This model is intended, among other functions, to facilitate the integration of AI-driven mechanisms with student-centered feedback in order to enhance the quality of learning. At this stage, the domain model is employed at a conceptual level to define key actors in the educational process and the relationships between them, to describe the feedback process within a course, and to structure assignment content and assessment criteria. The experiment presented in this study serves as a preparatory step toward the implementation and integration of the model into the educational process, highlighting its function as a conceptual framework for feedback design. Our results indicate that both types of feedback were generally perceived positively, but differences were observed in how their quality was evaluated. In one group, supervisor-provided feedback received higher ratings for clarity, depth, and relevance. At the same time, students in the other group showed a slight preference for feedback from ChatGPT 3.5, particularly in terms of improving their understanding of the assignment topics. The speed and consistency of AI-generated feedback were highlighted as key advantages, indicating its potential value in educational environments where personalized feedback from instructors is limited.

## KEYWORDS

AI, ChatGPT 3.5, domain model, experiment, personalized feedback

## 1 Introduction

Digitalisation has transformed the way students learn and teachers teach. It has enabled students to access educational resources from anywhere in the world, at any time, and at their own pace (Steriu and Stănescu, 2023). Automating the learning process may revolutionize learning and teaching by reducing teachers' workload and providing students with personalized learning experiences and feedback (Dhananjaya et al., 2024). Bill Gates says AI will replace doctors and teachers within 10 years and claims that humans will not be needed "for most things" (Zilber, 2025).

The presented work is part of our research program focused on developing a general and universal domain model for learning and teaching in higher education institutions, based on an abstract process domain model. This model should not only encompass process objectives and outcomes but also integrate a reliable feedback mechanism. The proposed domain model serves as a conceptual framework that defines and organizes the key concepts and their interrelationships related to the teaching and learning of a specific subject or topic in higher education. It provides a foundation for describing the learning content, including the set of skills, knowledge and strategies related to the tutored topic, expert knowledge, and potential misconceptions students may have. Additionally, it facilitates understanding and formalization of educational processes, including personalized monitoring and support for students as well as personalized feedback (Shvets et al., 2024).

One of the stages of this research program is the evaluation of the domain model for automated personalized feedback to students in the learning process. To evaluate our model, we utilize several methods:

- Evaluating the correctness of the model using real-life use case scenarios (Laskar et al., 2023; Nasiri et al., 2021; Suhail et al., 2022)
- Evaluating the domain model with the involvement of stakeholders to verify its alignment with the needs of users and its practical application in the educational process (Arora et al., 2024; Chowdhury et al., 2022; Owan et al., 2023)
- Conducting an experiment that simulates automated feedback according to the proposed domain model under real conditions within a course to evaluate the potential pedagogical effect of the automated feedback (Wang et al., 2023; Yu et al., 2021)

In the presented study, we focus on the last of the three points. We conduct an experiment in field conditions within a single academic course to test how our domain model can assist in automating personalized feedback for students. The aim of this experiment is to examine the impact of instant automated personalized feedback on students' performance. In one case, the feedback will be provided by the ChatGPT 3.5 chatbot. In the other case, the feedback will be given by a supervisor in the usual manner.

The structure of the paper is as follows: The research methodology is presented in section 2. In section 3, we present the results and conduct an analysis of students' performance and their satisfaction with the feedback. Finally, section 4 outlines the main findings, the study's limitations, and general recommendations for future research.

## 2 Materials and methods

### 2.1 The approach

The central element of any research is the selection of strategies, methods, or techniques that are applied to conduct the study (van Thiel, 2014). As part of our research, we decided to conduct an experiment in which we analyse students' satisfaction with the feedback received from both the teacher and ChatGPT 3.5 within the context of the learning course.

### 2.2 The procedure

The experiment involves 36 students from Taltech Virumaa College, with an average age of 30 years. The syllabus for the RAM0800 Digital Logic and Digital Systems study course (6 credits) in the Telematics bachelor's degree program was used for the experiment. The theoretical and practical materials of the course are studied during weeks 1 and 14 of the 17-week course. The experiment was conducted during the last 4 weeks of the course (weeks 14–17).

Students' tasks consist of four sets of assessment questions. The revised Bloom's taxonomy by Anderson et al. (2001) was used for the formulation and design of sets of assessment questions.

Each set of questions for each learning outcome is designed to include four questions: an A, a B, a C, and a D question. The A question covers the "Remember" level of Bloom's Taxonomy, the B question the "Understand" level, and the C question the "Apply" level. The D question covers the three higher levels (analysis, synthesis, and evaluation) of Bloom's taxonomy. The A questions are threshold questions necessary for passing the study course. Students achieving correct answers to questions A, B, and C receive an excellent grade. D questions are designed to indicate to students that there is still much to be learned. The number of attempts for solving each question is unlimited. The time for each attempt is limited depending on the complexity of the question.

As a mechanism for automated personalized feedback on these questions, we utilize the capabilities of ChatGPT 3.5. This innovative AI platform offers significant opportunities for enhancing personalized learning experiences and shaping the future of education (Albdrani and Al-Shargabi, 2023). ChatGPT 3.5 analyses students' responses and determines whether the correct answer has been provided, assigns grades, and offers feedback recommendations to the student.

While this feedback will be automated in the domain model we are developing, in the experiment, it was provided by the teacher, simulating the future automated feedback algorithm. To maintain a manageable workload for the teacher and simultaneously create high-quality research design, feedback using ChatGPT 3.5 was provided to only half of the students, while the other half received feedback in the usual manner. These groups alternated weekly to ensure fairness. Overall, the experiment consisted of four consecutive stages.

Stage 1: The students were divided into two groups (Group A and Group B, with 18 students each) using simple random sampling facilitated by a random number generator in Microsoft Excel. Due to the small sample size, stratified randomization was not employed. To ensure the equivalence of the groups at baseline, we conducted a comparative analysis of participants' key characteristics (age, knowledge level, academic performance) and confirmed that no statistically significant differences were present. This allowed us to consider the groups comparable prior to the start of the experiment.

Stage 2: All students were informed in advance that they would be participating in this experiment. Each student had the option to decline participation, in which case they could complete the course in the usual manner.

During the 14th and 15th weeks of the semester, the teacher provided the first and second sets of control questions (four questions in each set). The task for all students was to complete these control assignments. Upon completion, the task had to be uploaded as a text file to the Moodle platform. The response could be in the form of a

text-based essay or specific program code in the Falstad simulator.<sup>1</sup> Once a student uploaded their work to Moodle, the teacher received a notification.

Group A served as the focus group, receiving additional feedback. Additional feedback meant that students received a response from the teacher immediately after submitting their assignment. To generate this feedback, the teacher used ChatGPT 3.5. Group B acted as the control group and received feedback directly from the teacher within the standard timeframe (no later than 48 h after submission) during the same period.

The final deadline for completing the assignments was set at the end of the 15th week.

Stage 3: During the 16th and 17th weeks of the semester, the teacher provided the third and fourth sets of control questions (four questions in each set). At the beginning of week 16, Group B became the focus group, and Group A became the control group. The strict deadline for completing the tasks was set for the end of week 17.

Stage 4: At the end of the 15th and 17th weeks, a survey was conducted among the students. This survey gathered the students' opinions on the automated personalized feedback provided during the experiment.

## 2.3 Data collection procedures

### 2.3.1 Qualitative data collection

During the qualitative phase of this study, a carefully designed online survey was conducted using the Moodle survey tool. The aim of this survey was to assess the quality of feedback received by students from both the supervisor and ChatGPT 3.5. The survey questions underwent thorough review and refinement to ensure their clarity and comprehensiveness. Data collection took place over 2 weeks in the spring semester of 2024, resulting in 20 valid responses after Stage 2 and 23 valid responses after Stage 3. The difference in the number of valid responses across the two stages is explained by the absence or incompleteness of some questionnaires. There were no formal refusals to participate; the survey was voluntary and anonymous. However, some students did not complete the survey at one or both stages, while others submitted incomplete responses, which were excluded from the final analysis. Specifically, questionnaires with more than 20% of mandatory questions left unanswered, as well as those displaying inconsistent response patterns (e.g., identical ratings across all items without justification in open-ended fields), were excluded. Incomplete or incorrectly filled questionnaires were filtered out prior to analysis to ensure the reliability of the data.

Our survey is inspired by the work of Dawson et al. (2019) and Olsen and Hunnes (2024). However, in our study, we used only the abbreviated version of the survey, focusing on written feedback. Within the course, students primarily received written feedback through electronic annotations, i.e., typed comments in Moodle. Both Group A and Group B received feedback from both the supervisor and ChatGPT 3.5.

For data collection, the survey was conducted within the Moodle learning environment, and the data was exported and analyzed in MS Excel. The survey was conducted at the end of weeks 15 and 17 of the

academic term. The survey was personalized and conducted for all participants in the experiment (Supplementary Appendix A).

### 2.3.2 Quantitative data collection

The results of the sets of assessment questions completed by students were used as quantitative data.

We created a form for conducting the experiment and interacting with ChatGPT 3.5. The key feature of this form is that the AI must first generate its own solution and then compare it with the student's solution. Our task for ChatGPT 3.5 is to assess the student's solution. The form for conducting the experiment data is presented in Table 1.

After the experiment was completed, the results were exported to .xls files.

## 2.4 Data analysis procedures

After the qualitative data analysis, we intended to assess the quality of feedback. The analysis of students' responses to the questions evaluates the clarity, completeness, comprehensibility, and relevance of the feedback provided by the supervisor and ChatGPT 3.5 in the student survey conducted at the end of the experiment. First, we assessed the percentage of respondents for each question in the proposed survey. Second, we evaluated the average weight values of each question.

As a quantitative analysis, we conducted a comparison of the mean scores assigned by the supervisor and ChatGPT 3.5 for each assignment. We also compared the number of attempts made by students when feedback was provided by the supervisor and ChatGPT 3.5 for each assignment.

## 3 Results

### 3.1 Student evaluation of feedback

Students were asked to evaluate the feedback by indicating their level of agreement with six statements. Each response in the conducted survey has its own weight: 1—strongly disagree, 2—disagree, 3—neither disagree nor agree, 4—agree, 5—strongly agree, 6—not able to judge. In the first survey (after the 15th week), 56% of the total number of

TABLE 1 Form for conducting the experiment data.

Steps	The content of the used form
1	Question: question here
2	Student's solution: student's solution here
3	Actual solution: steps to work out the solution and your solution here (from teacher and from ChatGPT 3.5)
4	If the student's solution is the same as the actual solution just calculated: "yes" or "no"
5	Student grade: correct or almost correct or incorrect (if correct or almost correct then put points and give feedback, if incorrect the reject and give feedback)
6	Feedback: If extra feedback then generated by ChatGPT 3.5, else feedback from teacher.

<sup>1</sup> <https://www.falstad.com/circuit/>

students participated. In the second survey (after the 17th week), 64% of the students who took part in the experiment participated.

The results obtained after surveying Group A are presented in Figures 1, 2. Figures 3, 4 illustrate how students assessed the feedback in Group B after weeks 15 and 17, respectively. Group B was more satisfied on all aspects than Group A. This was the case regardless of whether they were rating the feedback received from the supervisor or from ChatGPT 3.5.

We compared the average satisfaction with feedback received from the supervisor and from ChatGPT 3.5. The result of this comparison is presented in Figure 5. Again, the differences between Group A and Group B are striking, with Group A (receiving ChatGPT 3.5 feedback in week 15 and supervisor feedback in week 17) being less satisfied than Group B. This was confirmed by a linear mixed models analysis with condition, week and group as fixed effects and student ID as random effect. While there was a trend toward a difference between groups A and B,  $F(1, 24.19) = 2.95$ ,  $p = 0.095$ , no effect of week,  $F(1, 20.87) = 0.086$ ,  $p = 0.773$  or condition,  $F(1, 20.87) = 0.42$ ,  $p = 0.523$ , could be discerned. Entering group as random factor instead of as fixed effect did not change the results for either week or condition. In other words, students seemed equally satisfied with both types of feedback.

### 3.2 Comparison of mean scores assigned by the supervisor and ChatGPT 3.5

We conducted a comparative analysis of the average scores obtained by students for the completion of control assignments as part of the experiment. Initially, all assignments were graded using a 100-point scale, which allowed for more precise differentiation of student performance. Subsequently, to ensure consistency and facilitate interpretation of the results, the scores were converted into the traditional five-point grading scale commonly used in educational practice.

We compared the average marks awarded to students depending on whether the assessment and feedback were provided by the human supervisor or ChatGPT 3.5. This approach enabled us to analyse the impact of different types of feedback on students' academic outcomes.

The comparison of average scores given by the supervisor and ChatGPT 3.5 is presented in Figure 6. Again, linear mixed-model analyses of differences between grades found a trend for group,  $F(1, 28.20) = 3.02$ ;  $p = 0.093$ , this time with an advantage for group A. However, no effect was found for either week,  $F(1, 27.63) = 0.07$ ,  $p = 0.79$ , or for condition,  $F(1, 27.63) = 0.02$ ,  $p = 0.83$ . Again, the same results for week and condition held when group was entered as a random instead of fixed effect. There was thus no evidence of any change in grades as a function of the kind of feedback.

### 3.3 Number of attempts made by students

We compared the number of student response attempts made to achieve the highest result. Given that each attempt rejected by the teacher was necessarily accompanied by a comment with personalized feedback for the student, this means that the number of attempts is directly related to the number of feedback comments.

The comparison of the average number of attempts made by students to complete tasks, when working with a supervisor and with ChatGPT 3.5, is presented in Figure 7. No difference between the two conditions was found,  $F(1, 27.55) = 1.23$ ,  $p = 0.277$ , nor of week,  $F(1, 27.55) = 1.23$ ,  $p = 0.277$ , nor of group,  $F(1, 28.10) = 0.001$ ,  $p = 0.97$ .

## 4 Discussion

The results of our study indicate that both supervisor-provided feedback and feedback from ChatGPT 3.5 were generally well received by students; no differences were observed in student satisfaction and

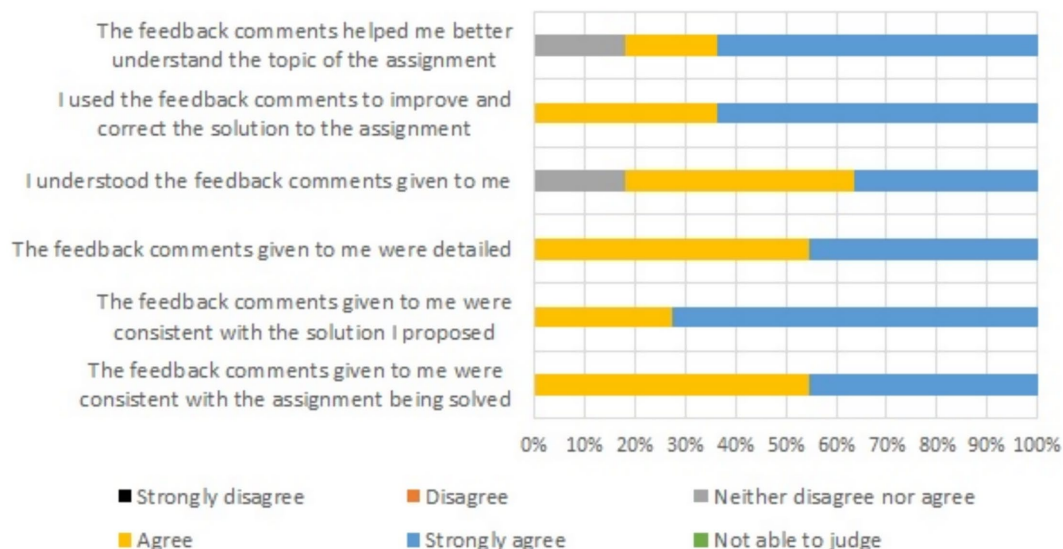


FIGURE 1

Degree of agreement regarding feedback from the ChatGPT 3.5 in the section of the course where written feedback was utilized. First stage (14–15 weeks). Group A. Percentage.



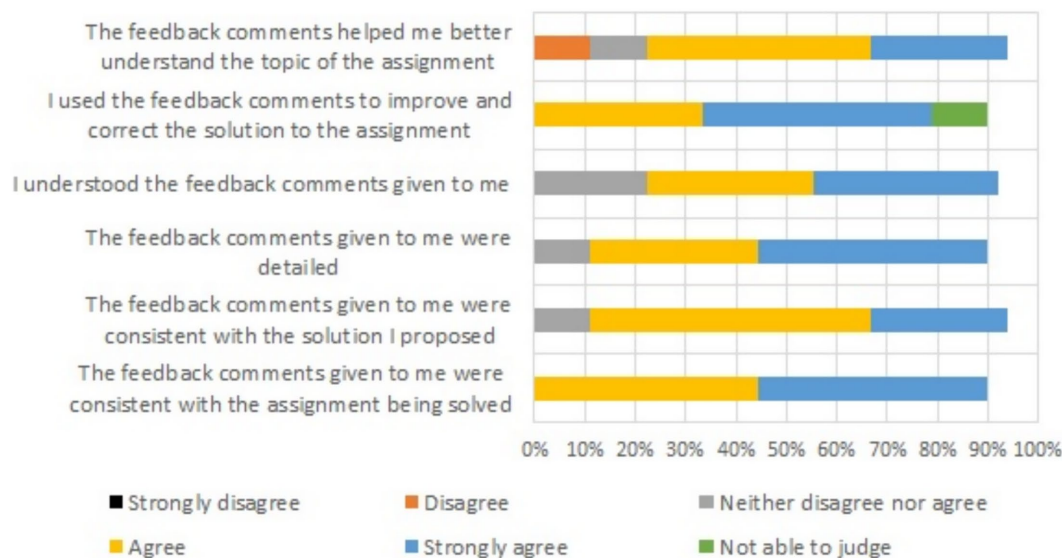


FIGURE 2

Degree of agreement regarding feedback from supervisor in the section of the course where written feedback was utilized. Second stage (16–17 weeks). Group A. Percentage.

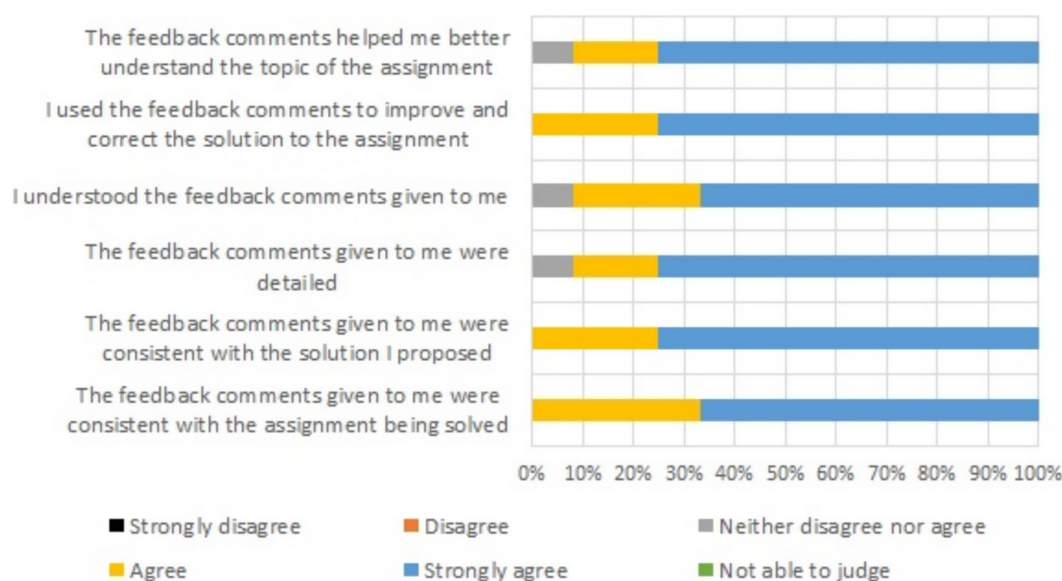


FIGURE 3

Degree of agreement regarding feedback from supervisor in the section of the course where written feedback was utilized. First stage (14–15 weeks). Group B. Percentage.

their perception of feedback quality. Overall, students from both groups (A and B) demonstrated a positive attitude toward both types of feedback, acknowledging their impact on improving their work, enhancing their understanding of assignment topics, and motivating them in their studies.

Regarding feedback quality, our findings show that students were generally equally satisfied with the clarity and level of detail in the comments provided by their supervisors.

A comparison of the average grades received by students based on the source of feedback (Figure 6) showed no differences in academic

performance. This suggests that in line with the similarity in the perceived quality of feedback, both types of comments led to similar academic outcomes. Furthermore, a comparison of the number of attempts made by students to achieve their best results (Figure 7) revealed a similar level of engagement with both types of feedback. This confirms that the effectiveness of feedback in encouraging students to revise and improve their work was comparable for both supervisor and ChatGPT 3.5 feedback.

Several limitations of this study warrant consideration. Firstly, the relatively small sample size ( $N = 36$ ) restricts the generalizability of the

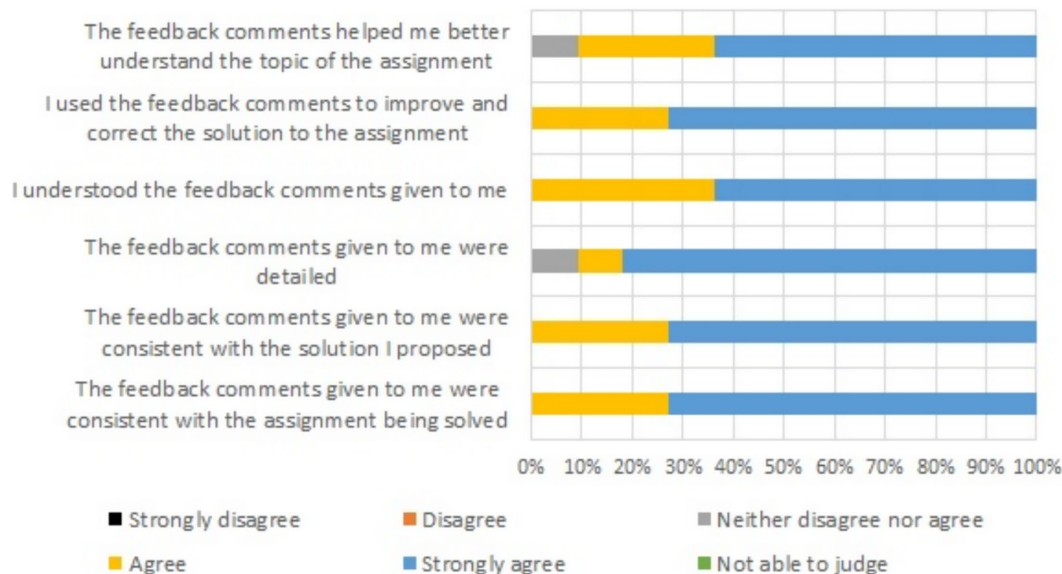


FIGURE 4

Degree of agreement regarding feedback from the ChatGPT 3.5 in the section of the course where written feedback was utilized. Second stage (16–17 weeks). Group B. Percentage.



FIGURE 5

Average satisfaction with feedback received from the supervisor and from ChatGPT 3.5.

findings. Although baseline equivalence between groups was confirmed, future research involving larger and more diverse samples would strengthen the reliability of the results and enhance their relevance across diverse educational contexts and disciplines.

Secondly, the AI-generated feedback was delivered exclusively via ChatGPT 3.5 and was not fully integrated with the domain model currently under development, nor supported by a specialized automated feedback system. Consequently, the feedback may have lacked the depth, adaptability, and contextual sensitivity that our model is intended to provide in the future.

Thirdly, the potential for experimenter bias warrants consideration. Although random allocation to groups was employed to minimize systematic bias and ensure representativeness, the supervisor's dual role—as both the provider of feedback in the control group and the assessor—may have inadvertently influenced students' perceptions or assessment outcomes. The involvement of independent

evaluators, as well as greater automation in the delivery of feedback, would likely mitigate this risk.

Finally, this study focused on short-term effects within a single course. Further research involving multiple courses and curricula is required to develop a more comprehensive understanding of the role that AI-generated feedback plays in supporting student learning and academic development.

The results of our study can be further contextualized within recent empirical research on the application of generative AI in higher education. For example, a randomized controlled trial conducted by [Noble and Wong Alan \(2025\)](#) at the University of Hong Kong investigated the effects of AI-generated feedback on students' written assignments, motivation, and emotional responses. Their study reported a significantly greater improvement in essay quality among students who received AI-generated feedback compared to those in the control group, underscoring the potential of such systems to enhance academic performance.

These findings align with our own and support the view that AI-generated feedback (e.g., via ChatGPT 3.5) can be as effective as feedback provided by human supervisors. This reinforces the broader hypothesis that generative AI tools can successfully complement or partially replace traditional feedback mechanisms, particularly under constraints of limited time and teaching resources. Our results further validate the pedagogical value of AI-driven feedback and its role in enhancing the human element of the learning process.

The experiment we conducted validates the functionality of the feedback component within our domain model for learning and teaching. This model enables an automated system to determine whether a student is prepared to begin a task and whether the outcome meets specified criteria. During task execution, the student's actions are aggregated into an actual result, which is then used to generate feedback. Although this architecture supports fully automated

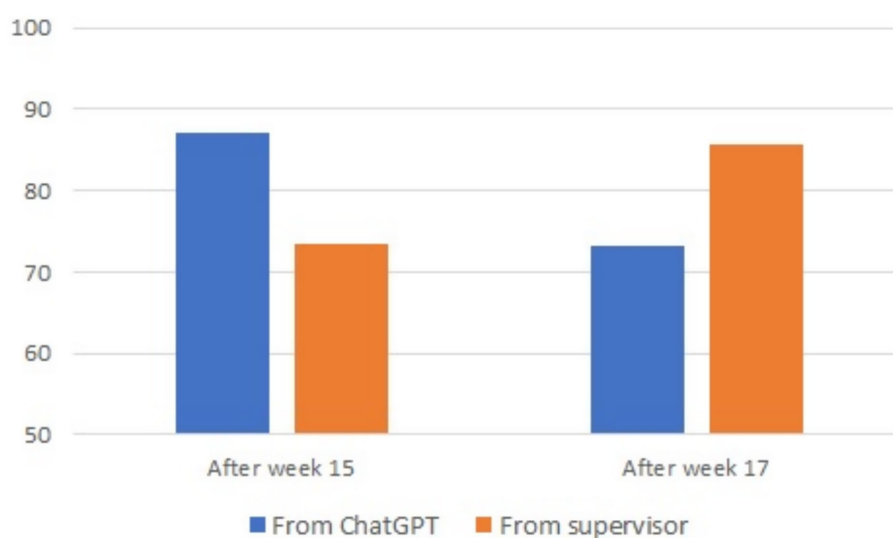


FIGURE 6  
Average grade on assignments from the supervisor and ChatGPT 3.5.

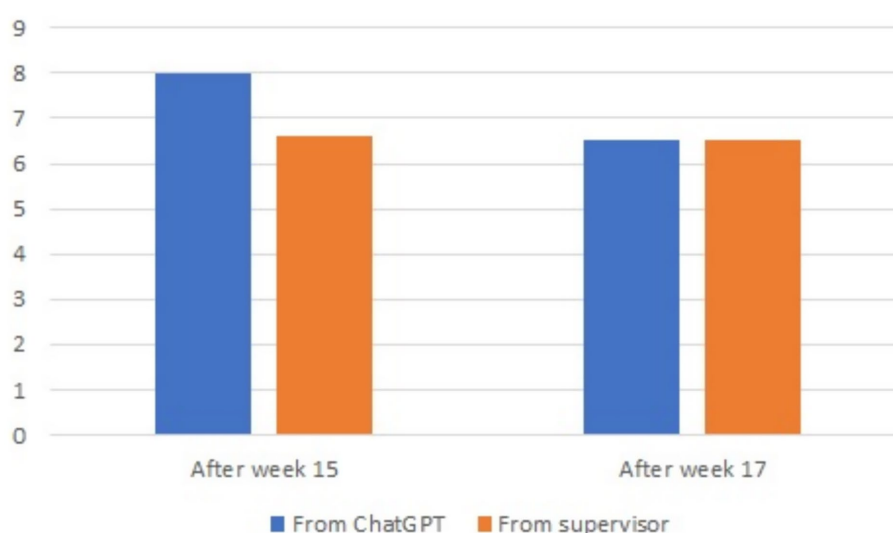


FIGURE 7  
Average number of attempts from the supervisor and ChatGPT 3.5.

feedback, in the present study the process was manually simulated with the involvement of a teacher.

Ultimately, our findings highlight the potential of AI-based feedback systems such as ChatGPT 3.5 as a valuable complement to traditional supervisor feedback, particularly in large-scale educational settings, where providing personalized comments to each student within a short timeframe can be challenging. However, the importance of supervisor feedback should not be underestimated, especially in situations requiring deep understanding and context-specific guidance.

Our findings inspire future research. The domain model for learning and teaching that we are developing should be universal, aimed at optimizing the combination of human and automated

feedback to enhance student engagement, learning outcomes, and overall satisfaction. Our experiment demonstrated that automated feedback is not inferior to manual feedback, which requires significantly more effort. This further underscores the importance of exploring processes for automating and personalizing feedback.

## Data availability statement

The original contributions presented in the study are included in the article/[Supplementary material](#), further inquiries can be directed to the corresponding author.

## Ethics statement

Ethical approval was not required for the studies involving humans because the study was conducted anonymously for the participants. All participants were informed that they were part of an experiment. Each participant had the option to withdraw from the experiment. The presented results do not include any personal data of the participants. The studies were conducted in accordance with the local legislation and institutional requirements. Written informed consent for participation was not required from the participants or the participants' legal guardians/next of kin in accordance with the national legislation and institutional requirements.

## Author contributions

OS: Writing – review & editing. KM: Writing – review & editing. GP: Funding acquisition, Writing – review & editing. MM: Writing – review & editing.

## Funding

The author(s) declare that no financial support was received for the research and/or publication of this article.

## References

- Albdrani, R. N., and Al-Shargabi, A. A. (2023). Investigating the effectiveness of ChatGPT for providing personalized learning experience: a case study. *Int. J. Adv. Comput. Sci. Appl.* 14, 1208–1213. doi: 10.14569/IJACSA.2023.01411122
- Anderson, L. W., Krathwohl, D. R., and Bloom, B. S. (2001). A taxonomy for learning, teaching, and assessing: a revision of Bloom's taxonomy of educational objectives. New York: Addison Wesley Longman, Inc.
- Arora, C., Grundy, J., and Abdelrazek, M. (2024). "Advancing requirements engineering through generative AI: assessing the role of LLMs" in Generative AI for effective software development (Springer), 129–148. doi: 10.1007/978-3-031-55642-5\_6
- Chowdhury, N., Katsikas, S., and Gkioulos, V. (2022). Modeling effective cybersecurity training frameworks: a Delphi method-based study. *Comput. Secur.* 113:102551. doi: 10.1016/j.cose.2021.102551
- Dawson, P., Henderson, M., Mahoney, P., Phillips, M., Ryan, T., Boud, D., et al. (2019). What makes for effective feedback: staff and student perspectives. *Assess. Eval. High. Educ.* 44, 25–36. doi: 10.1080/02602938.2018.1467877
- Dhananjaya, G. M., Goudar, R. H., Kulkarni, A. A., Rathod, V. N., and Hukkeri, G. S. (2024). A digital recommendation system for personalized learning to enhance online education: a review. *IEEE Access* 12, 34019–34041. doi: 10.1109/ACCESS.2024.3369901
- Laskar, M. T. R., Bari, M. S., Rahman, M., Bhuiyan, M. A. H., Joty, S., and Huang, J. X. (2023). "A systematic study and comprehensive evaluation of ChatGPT on benchmark datasets" in Proceedings of the Annual Meeting of the Association for Computational Linguistics, 431–469. doi: 10.48550/arXiv.2305.18486
- Nasiri, S., Rhazali, Y., Lahmer, M., and Adadi, A. (2021). From user stories to UML diagrams driven by ontological and production model. *Int. J. Adv. Comput. Sci. Appl.* 12, 333–340. doi: 10.14569/IJACSA.2021.0120637
- Noble, L., and Wong Alan, C. S. (2025). The impact of generative AI on essay revisions and student engagement. *Comput. Educ. Open*:100249. doi: 10.1016/j.caeo.2025.100249
- Olsen, T., and Hunnes, J. (2024). Improving students' learning—the role of formative feedback: experiences from a crash course for business students in academic writing. *Assess. Eval. High. Educ.* 49, 129–141. doi: 10.1080/02602938.2023.2187744
- Owan, V. J., Abang, K. B., Idika, D. O., Etta, E. O., and Bassey, B. A. (2023). Exploring the potential of artificial intelligence tools in educational measurement and assessment. *Eurasia J. Math. Sci. Technol. Educ* 19:em2307. doi: 10.29333/ejmste/13428
- Shvets, O., Murtazin, K., Meeter, M., and Pihö, G. (2024). Towards a domain model for learning and teaching. In: International Conference on Model-Driven Engineering and Software Development. pp. 288–296. doi: 10.5220/0012471400003645
- Steriu, I., and Stănescu, A. (2023). Digitalization in education: navigating the future of learning. In: Proceedings of the International Conference on Virtual Learning, pp. 169–182. doi: 10.58503/icvl-v18y202314
- Suhail, S., Malik, S. U. R., Jurdak, R., Hussain, R., Matulevičius, R., and Svetinovic, D. (2022). Towards situational aware cyber-physical systems: a security-enhancing use case of blockchain-based digital twins. *Comput. Ind.* 141:103699. doi: 10.1016/j.compind.2022.103699
- van Thiel, S. (2014). Research methods in public administration and public management: an introduction. London: Routledge. doi: 10.4324/9780203078525
- Wang, H., Tlili, A., Huang, R., Cai, Z., Li, M., Cheng, Z., et al. (2023). Examining the applications of intelligent tutoring systems in real educational contexts: a systematic literature review from the social experiment perspective. *Educ. Inf. Technol.* 28, 9113–9148. doi: 10.1007/s10639-022-11555-x
- Yu, S. J., Hsueh, Y. L., Sun, J. C. Y., and Liu, H. Z. (2021). Developing an intelligent virtual reality interactive system based on the ADDIE model for learning pour-over coffee brewing. *Comput. Educ. Artif. Intell.* 2:100030. doi: 10.1016/j.caeai.2021.100030
- Zilber, A., (2025). No title [WWW document]. New York Times Available online at: <https://nypost.com/2025/03/27/business/bill-gates-said-ai-will-replace-doctors-teachers-within-10-years/>

## Conflict of interest

The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

## Generative AI statement

The authors declare that Gen AI was used in the creation of this manuscript for conducting the experiment. ChatGPT 3.5 was part of the experiment conducted.

## Publisher's note

All claims expressed in this article are solely those of the authors and do not necessarily represent those of their affiliated organizations, or those of the publisher, the editors and the reviewers. Any product that may be evaluated in this article, or claim that may be made by its manufacturer, is not guaranteed or endorsed by the publisher.

## Supplementary material

The Supplementary material for this article can be found online at: <https://www.frontiersin.org/articles/10.3389/feduc.2025.1624516/full#supplementary-material>