Gamification Powered by a Large Language Model to Enhance Flipped Classroom Learning in Undergraduate Computer Science

Kathleen Kelly, Bo Wu, and Christine Liebe
Colorado School of Mines, Golden, CO, USA

kathleenmariekelly@mines.edu
bwu@mines.edu
cliebe@mines.edu

Abstract: The flipped classroom in computer science has become a popular pedagogy where students complete preparatory reading and videos before class, leaving class time for hands-on practice. In theory, students come to class predisposed with knowledge to complete programming tasks, quizzes, and homework assignments. However, if students do not complete or do not fully understand the preparatory material, those students will not be equipped for class and may also be embarrassed to ask questions during class. In this study, we created a gamified quiz that is powered by a Large Language Model (LLM) that loads in questions from the flipped content transcripts. The game offers students a fun assessment after completing the flipped content that will both motivate them and prepare them for class. The game also provides an anonymous way for students to ask questions as they progress through the game content. Those questions can then be analysed by the instructor and reviewed in class. The study compares in-class quiz scores of a control group versus a group that had access to the gamified quiz in an undergraduate software engineering course over three consecutive course modules. The results obtained from both quantitative and qualitative data are promising, showing an increase in student in-class quiz scores along with positive student engagement with the gamified quiz. Students also reported that the game made them more comfortable with asking questions. Future work would extend this research to more students and over a longer period. The LLM and gamification also have much potential, where games can be created dynamically and be customized and personalized for each student's needs. This work shows an exciting gap in research that could lead to gamified experiences for students in the flipped model that are both relevant and personalized with the use of a LLM, offering the students the greatest chance for success.

Keywords: Gamification, Computer Science Education, Engagement, Satisfaction, Games, Feedback, Large Language Models, Flipped Classroom, Gamified Learning

1. Introduction

The flipped classroom model is a popular teaching method where students are required to complete preparation activities before coming to class, such as watching lecture videos or reading relevant content. By doing this work in advance, the intent is that students are better prepared to come to class with more specific questions and more understanding of the content, ready to perform hands-on tasks. However, the flipped model also creates challenges for instructors whereby students may not be fully grasping the flipped content, yet do not ask questions during class. Students often skip the flipped content due to time constraints or lack of motivation, as stated by Nielsen (2023). Assessing the students' understanding with quizzes can result in poor grades, leaving the instructor and students feeling discouraged.

Alahmari et al. (2023) define educational gamification as the idea of adding game elements to a course environment to help motivate and engage students in a playful manner while also teaching the content. This pedagogy is a teaching tool yet has its own challenges of striking the right balance between fun and learning. Creating an educational game that can be modified with different questions depending on the course module is time consuming for the instructor. Both flipped classrooms and educational games are making use of Large Language Models (LLMs) to provide a more personalized learning experience and increase student motivation and performance. LLMs are becoming a strategy for saving time for the instructor and for keeping content relevant and up to date.

In this study, the use of a game powered by a LLM was studied in conjunction with a flipped classroom. Specifically, the following research questions (RQs) were investigated:

RQ1: Does preparatory flipped content reinforced with a LLM-powered gamified quiz increase student learning outcomes?

RQ2: Does classroom material enhanced with gamification increase student motivation?

RQ3: Does the allowance of asking anonymous questions on a gamified quiz increase student comfort level?

2. Literature Review

2.1 The Flipped Classroom and Computer Science Education

The flipped classroom model increases student performance and high order thinking. Zheng et al. (2020) stated that the flipped model had "significant impact" on the motivation of students. Subramaniam et al. (2019) suggested that in the flipped model, the teacher is the "guide on the side" as opposed to the "sage on the stage". Jdaitawi (2019) found that students in a flipped classroom have a higher level of self-regulation and social connectedness.

In a different study, Nielsen (2023) looked at the frustrations that students experience in a flipped classroom. Nielsen found that some students lacked motivation to do preparatory work, leading to unpreparedness in class. Urquiza-Fuentes (2023) suggested that a motivating factor for students to complete the flipped content was to use rewards, such as assessments that contribute towards a final grade. The common theme among the studies was that the flipped classroom model had the potential to help student engagement and self-regulation, but that there is "no one size fits all" model, as noted by Subramaniam et al. (2019). Hwang et al. (2019) commented that it is not a question of whether the flipped model works, but how to make it more effective.

To address some of the student frustrations, such as student questions that arise while completing the preparatory work, Uchiyama et al. (2023) proposed a system that added a LLM to give immediate answers to students. However, the uncertainty of the LLM responses could lead to more work for the instructor, having to filter through the LLM-student conversations and correct any errors or misconceptions. Huang et al. (2023) looked at student motivation in the flipped classroom and the use of a personalized recommendation system powered by a LLM as part of a flipped classroom model.

The flipped classroom model has been shown to be an effective pedagogy in computer science classrooms while also having challenges. By leveraging LLMs with the flipped classroom, there is the potential for more personalization for students as well as immediate feedback to student questions.

2.2 Gamification and Computer Science Education

Gamification is the idea of adding game elements to an area that does not otherwise relate to games, as stated by Alahmari et al. (2023). In one study, Khaldi et al. (2023) revealed that the trio of points, badges, and leaderboards (PBL) are used the most. Alahmari et al. (2023) looked at the impact of gamification on student learning, concluding that gamification makes the learning experience more enjoyable, helps with engagement in social situations, and is good for psychological skills. Sadiku et al. (2023) conducted research on gamification in computer science with an angle towards competition. Cao (2023) researched using gamification and a LLM to help students learn programming concepts and increase their sense of belonging. Lanzi et al. (2023) used a LLM for collaborative game design to simulate the process performed by human game designers.

Like the flipped classroom model, gamification has the potential to be more effective with the addition of personalization for each student. Leveraging LLMs, educational games could be created with personalized feedback and personalized design to better motivate students.

2.3 Gamification and the Flipped Classroom

Combining gamification with the flipped classroom provides a modern way of engaging students. In a review of flipped learning with gamification, Ekici (2021) shared that students learn at their own pace using both intrinsic and extrinsic motivation. Ekici (2021) suggested combining the flipped classroom with gamification, "bringing the best of both worlds together to promote learning".

Gamification combined with the flipped classroom is a new theoretical framework being used and one that still needs more research. The game must be well designed and be a source of enhancement to the flipped material but not incur a novelty effect that could wear off, as stated by Ekici (2021).

2.4 LLMs and Computer Science Education

LLMs are being used in the context of computer science education as well as in the context of gamification. One motivation for using LLMs is to reduce instructor load while creating educational content. Song et al. (2024) created a novel LLM with prompt chains to generate multiple-choice questions according to a course outline, showing promising results in the quality of the questions. Tran et al. (2023) analysed LLM capabilities around

generating high-quality multiple-choice questions and answers for computing. The results showed accuracy, but that human intervention was still needed to vet the output of the LLM. Meißner et al. (2024) introduced a novel EvalQuiz to generate self-assessment quizzes via a LLM. The results showed that the auto-generated quizzes covered the appropriate material and saved instructor time, but still needed better prompt engineering to achieve "originality and versatility" between the questions.

While the flipped classroom, gamification, and the use of LLMs are each being used as pedagogical tools, combining the three techniques together is an area in need of research and one that has a large gap. This research will use a LLM to generate multiple-choice questions for a gamified quiz using transcripts from the flipped content videos. The use of the gamified quiz will then be used in this study to answer the proposed research questions.

3. Methods and Experiment

In this study, the flipped classroom model was already being utilized in three sections of a college software engineering course. The sections included 161 undergraduate sophomore-level students. This author was the instructor for all three sections, giving the students flipped content videos to watch before each class, followed by pop quizzes sporadically in class. Before this study, results showed that students were not performing well on the pop quizzes, but also not asking questions in class even when encouraged.

To try to better understand how students were performing in the flipped model, this project used a gamified quiz to add motivation and help with knowledge retention. The quiz was loaded with questions created by a LLM to make it more scalable across modules and require less instructor interaction. With this balance of fun and learning, this study looked at how adding a gamified quiz to a flipped model could help with student performance on pop quizzes.

3.1 Project Description

GameMaker Studio YoYo Games Ltd (2023) was used to create the novel game, "Engin-AIR-ing", where the player would answer multiple-choice questions related to the flipped content. While reading the question and answers, a clock would be running to encourage speed and leaves would be floating to the ground to encourage game play. The player would try to select the correct answer while also keeping leaves afloat to earn more points. To make the game accessible from any computer, it was built as an HTML5 package, then launched via a flask server written in python.

The flask server connected to GPT-3.5 Turbo using an OpenAI (2023) key. Using the python library from Liu (2022), llama_index, the transcripts from the flipped content videos were loaded into the LLM query engine. Once all the documents were loaded, a detailed prompt was sent to the query engine as follows:

"Create at least 10 unique multiple-choice questions with 4 different answers from the content. Be sure the questions are unique and cover all the content. There should be a variety of topics in the questions. Start each question with the word Question. One answer should be correct, the other 3 should be incorrect. Separate the question and answers with a newline. End the line with the correct answer with the word Correct in parentheses, like (Correct). Then give an explanation sentence for why the correct answer is correct. Return your response in json format. The json for each question should have fields for the question text, array of answers, explanation text, and correct answer text, like: \{"question": ,"answers": [ ],"explanation": ,"correct_answer": \} The final json will be an array named "questions" of these structures, one for each question. Do not put the word "Correct" after any answer. Be sure and include questions that span all the content."

The LLM json response was stored in memory until requested by the game via an HTML call. To ensure the questions were relevant, accurate, and not too similar to those that would be presented on the pop quiz in class, the instructor reviewed the questions and re-executed the LLM prompt until a suitable set was created.

Once the game had access to the quiz questions, it would begin displaying them one at a time, starting a timer, and sending leaves floating. The student would move the game player back and forth along the bottom of the screen using arrow keys. The spacebar key would cause the game player to blow air at a leaf above, keeping it afloat. The question and answers were displayed at the top of the screen where choosing an answer involved using the number keys.

As the player completed a question, a detailed explanation would be displayed along with a textbox area for the student to ask any questions. Data, including the number of correct answers, incorrect answers, leaves that

were kept afloat, and time on the question, was sent back to the flask server to be stored in a file. Any questions typed into the textbox were also sent back to the flask server to be stored for the instructor. Once the student completed all ten questions, the final score would be displayed in a leaderboard with any other students that had played. See Figure 1 for a screenshot of the game.

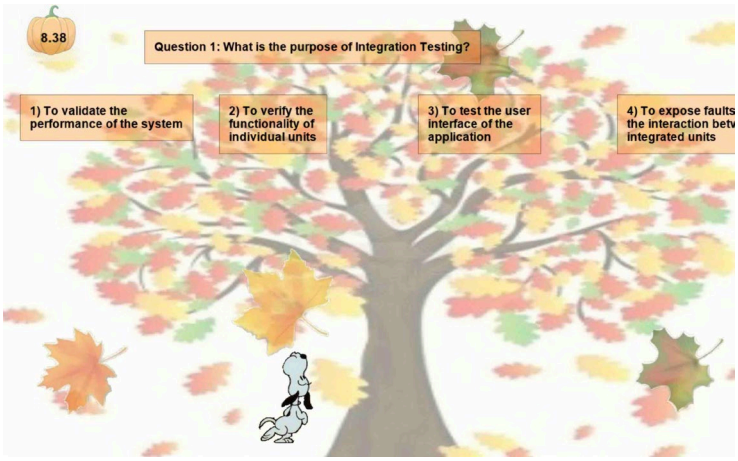

Figure 1: Engin-AIR-ing game, tenor (2015)

3.2 Experiment

Once the game was completed, the research experiment began with volunteer students from the software engineering course. Of the 161 students enrolled in the course, 23 students chose to be part of the project and signed informed consents. From the 23 students, 13 were chosen at random to receive access to the game while the other 10 students were the control group, not receiving access to the game.

The experiment ran over three lesson modules. The students were given access to the usual flipped content videos to prepare for class. The recipients of the game were also given access to play the game, being quizzed in a gamified way on the content from the videos. In class, all students were given a non-graded quiz. At the end of the experiment, students were administered a post-survey with questions about the flipped model and the game.

Quantitative data collected included in-class quiz scores, game scores, number of game attempts, number of correct and incorrect answers to each game question, number of leaves (floaters) left after each game question, and amount of time spent on each question. Statistical pre-condition tests were performed to determine if the data was normally distributed and had non-significant differences in variances, followed by the use of a two-sample t-test. Quantitative data was also analysed manually looking for trends in the data, such as how individual student's game scores or times improved across the three modules. Qualitative data collected in the post-survey offered questions with a likert scale as well as open-ended feedback questions.

4. Results and Discussion

4.1 RQ1: Does Preparatory Flipped Content Reinforced with a LLM-powered Gamified Quiz Increase Student Learning Outcomes?

After each of three modules, the 23 participating students were given an in-class ungraded quiz. The hypothesis was that the 13 students who had watched the flipped content and played the gamified quiz would score higher than the 10 students who had only watched the flipped content. After completion of the first two modules, students who had access to the flipped content and the gamified quiz scored higher than students who only had access to the flipped content. However, the third module had the opposite results, as shown in Figure 2.

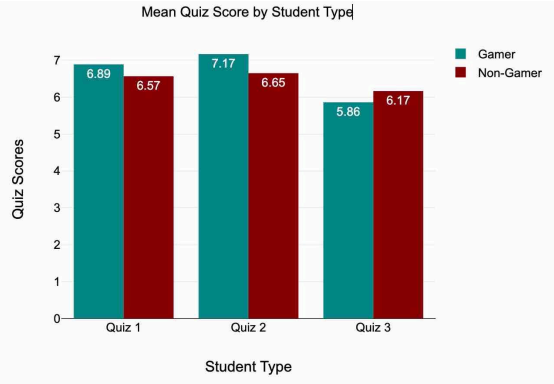

Figure 2: Mean Quiz Score by Student Type

This unexpected result from the third module could be due to the small sample size. This result could also have been related to the dwindling number of students per-module. By module three, the number of students playing the gamified quiz had decreased 30%.

To further analyse the results, statistical tests were performed on the two groups of students: the control group and the group that had access to the gamified quiz. First, a Shapiro-Wilk test was used to ensure the data was normally distributed. Running this test on all of the quiz data (N=65) did not show a significant departure from normality, $W(65) = 0.98$, $p = 0.396$. Second, a Levene test was performed to compare the variances in the data. This test concluded that the difference between the sample variances was not big enough to be statistically significant.

Because the data was found to be normally distributed and with similar variances, a two-sample t-test was performed to compare the means of the two groups in hopes of supporting the research hypothesis. Using a t-test calculator, the t-score and p-value for each of the three modules was determined, as shown in Table 1. In each case, however, the p-value was not less than the significance level of 0.05, meaning there was not enough evidence to show that the mean quiz scores between the two groups were different. This also could be due to the small sample size of students and the drop in motivation by the third module.

Table 1: Two Sample t-test

Module	t score	Degrees of freedom	p value
1	0.52	20	0.31
2	0.74	20	0.23
3	-0.39	20	0.35
1-3	0.43	64	0.33

The experiment running over three classes led to less participation as time went on, possibly due to busy student schedules and/or a lack of motivation over an ungraded in-class quiz. The participation level dropped in each class, possibly due to overall semester burnout or the novelty of the experiment wearing off.

4.2 RQ2: Does Classroom Material Enhanced With Gamification Increase Student Motivation?

To measure motivation, both quantitative and qualitative data were used. While playing the game, the study collected the number of floaters that were remaining after each question was answered, how long each question took to answer, and how many attempts of the gamified quiz were made. The post-survey also asked some questions related to student motivation with and without the gamified quiz.

Game play included having to keep leaves in the air while reading and answering multiple-choice questions, adding points to the total score. The leaves did not affect the correct answer in any way, they were purely a game play element. Results showed an upward trend of floaters across all 30 quiz questions, possibly a result of student motivation to interact with the game.

Looking at students who participated in the gamified quiz for all three modules, game scores and times were analysed. Specifically, three players consistently played all three module quizzes, resulting in total quiz scores and total time spent on the quizzes. Figure 3 shows that students who completed all three modules earned

better overall game scores across the three modules, improving their interaction with the game. This could be a measure of student motivation, showing the students tried harder on each attempt with the game.

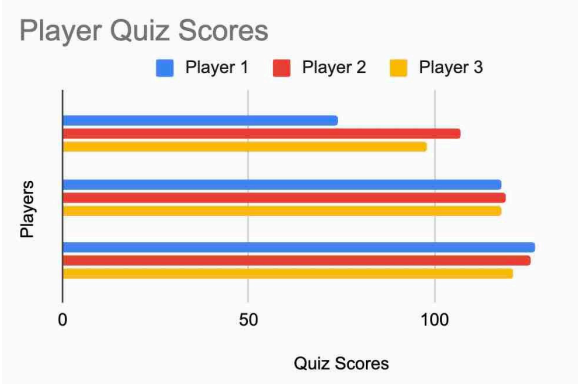

Figure 3: Player Quiz Scores Across Three Modules

Similarly, Figure 4 shows that students' time with the game decreased across the three modules, possibly showing their improved interaction with the game as they were able to answer questions correctly in less time as they continued to play. There were also students who replayed the game multiple times. Replaying the game resulted in the student being presented with the same game questions in the same order. The multiple game play attempts showed another form of motivation: students wanting to play the game more than once simply to improve their total score, total time, and leaderboard ranking. This is an exciting result, showing students are motivated by game elements of points and a leaderboard.

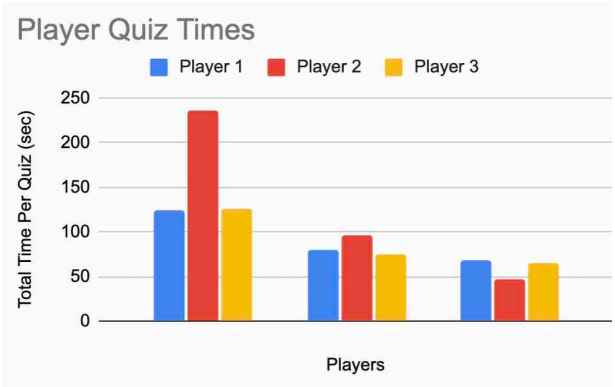

Figure 4: Player Quiz Times Across Three Modules

Qualitative data was collected from a post-survey that included questions related to student motivation, the flipped model, and its effectiveness. When asked which parts of the flipped model were most effective, students admitted to liking the flexibility of the flipped model, the short duration of videos, and the pop quizzes in class to reinforce their learning. When asked which parts of the flipped model were the least effective, students commented that having to do preparatory work was time consuming and many times they did not watch the videos due to lack of engagement.

When asked about how the game made a difference in learning, there were seven student responses. Overall, they found the game motivating, reinforcing, and enjoyable. When asked about its effectiveness, students commented that they liked the number of questions, they liked that the game immediately reinforced the video content, they liked the leaderboard, and they liked the timer as a speed element. When asked about ineffectiveness, the timer and leaves were mentioned as causing stress or making it hard to concentrate.

Specifically, some of the student comments included:

- *"I liked the time aspect of it, my brain was forced to recall the information quicker and that helped."*
- *"The leaderboard, it made me want to keep playing, to beat the other players' scores."*
- *"I think that this game is a step towards a better flipped classroom."*
- *"I thought I should note that I didn't get to play the game one of the times before class and I noticed I got a much lower score on the quiz."*

Overall, these results showed student motivation to play a gamified quiz, but not necessarily a quiz enhanced by a LLM. Future work would include adding more levels to the game where difficulty of the LLM questions would increase. Levels could also be personalized, allowing students to choose the content they would most like to see in the game.

4.3 RQ3: Does the Allowance of Asking Anonymous Questions on a Gamified Quiz Increase Student Comfort Level?

The gamified quiz provided a way for students to submit anonymous questions at the end of each quiz question. These student questions were saved into a text file for the study, the intent being that they could be used by the instructor in class. In this study, there was only one question asked over the duration of the three modules. Although this is a small number, it aligns with the current number of questions asked in class. The post-survey showed 100% agreement that an anonymous format for questions made students more comfortable. It is likely that students prefer asking a question in an anonymous way, but this could be provided outside of a gamified quiz. As LLMs become more reliable, future work would include a way to interact with the LLM in the game, being able to ask the LLM a question and receive an immediate answer.

5. Conclusion

The results of this experiment are promising, showing that students who played the gamified quiz performed better on pop quizzes and were motivated to play the game. Using a LLM as the source of the game questions is a newly adopted approach, allowing for the game to be filled with content from any source. Student feedback shows exciting promise for future work around adding gamification to the flipped classroom model.

We do acknowledge several limitations in this research project. First, the sample size of participants was small, only 23 total students out of a class of 161 students volunteered to be a part of the study. There could have been an internal validity threat of selection whereby those 23 students were better overall performers in the class and had the extra time for a research study. Second, the game itself could have been overwhelming for non-visual learners or learners that do not read quickly.

Third, the game used a LLM in a limited capacity due to the LLM being inconsistent with its responses. We found that even when giving the LLM a directory of content to use as its question bank, it tended to load the bank with very similar questions from one of the transcript files instead of across all transcript files. The questions generated by the LLM had to be vetted as they tended to contain duplicates, focus too much on one concept of the transcript, or were too similar to the pop quiz questions, which would give the game students a huge advantage.

Fourth, the experiment was short in duration due to the semester time constraints. The in-class pop quiz was ungraded to be fair to the entire class. This could have resulted in less reliable student scores as students were made aware of the ungraded pop quiz. The game itself was created in a very limited time frame so was constrained to only one level and a simple user interface, which could have led to the loss of interest by students.

There are many future directions for this project. Using a LLM-powered game to supplement flipped content in computer science education is a novel approach. This project showed the promise of this idea with its small sample size and short duration. Future work could take this project across an entire semester of a course using the entire student population in that course. Future work would also involve a much more complicated game engine that would utilize a LLM in more powerful ways, allowing more personalization of the game for each individual student. In doing so, student performance and engagement could be measured with a game containing more hands-on activities that better align with the flipped classroom pedagogy. Lastly, future work could analyse three groups: 1) students who only watch the flipped content, 2) students who watch the flipped content and do a simple practice assessment, 3) students who watch the flipped content and play a gamified quiz. This would add the dimension of simple assessment versus gamified assessment to the research.

The flipped model pedagogy has many benefits for students, giving them more time in class to put their knowledge into practical applications. However, students must be motivated to complete the preparatory flipped content, or the pedagogy fails for everyone. The flipped content must also remain relevant and have the right balance of engagement and learning. Gamification is a strategy for adding game elements to a classroom, but also must strike the right balance that meets the needs of any student. This study looked at the novel idea of adding a gamified quiz of content provided by a LLM to supplement the flipped material. The results of the study showed increased assessment scores in class and student motivation to play the gamified quiz. Future

directions for this could leverage the LLM in more powerful ways, leading to gamified experiences for students in the flipped model that are both relevant and personalized. This would offer students the greatest chance for success both in and out of the classroom.

References

(2015). *Snoopy GIF,* tenor, https://tenor.com/search/snoopy-autumn-gifs.

(2023). https://chat.openai.com.

Alahmari, M., Jdaitawi, M. T., Rasheed, A., Abduljawad, R., Hussein, E., Alzahrani, M., and Awad, N. (2023). "Trends and gaps in empirical research on gamification in science education: A systematic review of the literature", *Contemporary Educational Technology,* 15(3), ep431. Doi: 10.30935/cedtech/13177.

Cao, Chen (2023). "Scaffolding CS1 Courses with a Large Language Model-Powered Intelligent Tutoring System", *Association for Computing Machinery*, pp. 229–232. Doi: 10.1145/3581754.3584111.

Ekici, M. (2021) "A systematic review of the use of gamification in flipped learning", *Education and Information Technologies*, 26(3), pp. 3327-3346. doi: 10.1007/s10639-020-10394-y.

Huang, Anna Y.Q., Lu, Owen H.T., and Yang, Stephen J.H. (2023). "Effects of artificial Intelligence–Enabled personalized recommendations on learners' learning engagement, motivation, and outcomes in a flipped classroom", *Computers & Education,* 194, pp. 104684. doi: 10.1016/j.compedu.2022.104684.

Hwang, G.J., Yin, C. and Chu, H.C. (2019). "The era of flipped learning: promoting active learning and higher order thinking with innovative flipped learning strategies and supporting systems", *Interactive Learning Environments,* 27(8), pp. 991-994. Doi: 10.1080/10494820.2019.1667150.

Jdaitawi, Malek Turki (2019). "The Effect of Flipped Classroom Strategy on Students Learning Outcomes", *International Journal of Instruction*. https://api.semanticscholar.org/CorpusID:162176750.

Khaldi, A., Bouzidi R., and Nader F. (2023). "Gamification of e-learning in higher education: a systematic literature review", *Smart Learning Environments*, 10(1), pp. 10. doi: 10.1186/s40561-023-00227-z.

Lanzi, Pier Luca and Loiacono, Daniele (2023). "ChatGPT and Other Large Language Models as Evolutionary Engines for Online Interactive Collaborative Game Design", *Association for Computing Machinery,* pp. 1383-1390. doi: 10.1145/3583131.3590351.

Liu, Jerry (2022.) LlamaIndex*,* 10.5281/zenodo.1234, https://github.com/jerryjliu/llama_index.

Meißner, Niklas, Speth, Sandro, Kieslinger, Julian, and Becker, Steffen (2024). "EvalQuiz – LLM-based Automated Generation of Self-Assessment Quizzes in Software Engineering Education", *Gesellschaft für Informatik e.V.*, pp, 53–64. doi: 10.18420/seuh2024_04.

Nielsen, Kjetil Liestøl (2023). "Why Can the Flipped Classroom Frustrate Students? Experiences from an Engineering Mathematics Course", *Education Sciences,* 13(4), pp. 396. doi: 10.3390/educsci13040396.

Sadiku, Matthew N. O, Chukwu, Uwakwe C., and Sadiku, Janet O. (2023). "Gamification in Computer Science", *Central Asian Journal of Mathematical Theory and Computer Sciences,* 4(5).

Song, Tian, Tian, Qinqin, Xiao, Yijia, and Liu, Shuting (2024). "Automatic Generation of Multiple-Choice Questions for CS0 and CS1 Curricula Using Large Language Models", *In Computer Science and Education. Computer Science and Technology. ICCSE 2023*, Singapore, pp. 314–324. doi: 10.1007/978-981-97-0730-0_28.

Subramaniam, Muniandy Balakrishnan and Rani, Suwarna (2019). "The Effect of Flipped Classroom on Students' Engagement". *Technology, Knowledge and Learning,* 24(3), pp. 355-372. doi: 10.1007/s10758-017-9343-y.

Tran, Andrew, Angelikas, Kenneth, Rama, Egi, Okechukwu, Chiku, Smith, David H., and MacNeil, Stephen (2023). "Generating Multiple Choice Questions for Computing Courses Using Large Language Models", *2023 IEEE Frontiers in Education Conference (FIE),* College Station, TX, USA, pp. 1–8. doi: 10.1109/FIE58773.2023.10342898.

Uchiyama, Shintaro, Umemura, Kyoji, and Morita, Yusuke (2023). "Large Language Model-based System to Provide Immediate Feedback to Students in Flipped Classroom Preparation Learning". doi: 10.48550/arXiv.2307.11388.

Urquiza-Fuentes, Jaime (2023). "A Study About the Impact of Rewards on Student's Engagement with the Flipped Classroom Methodology", *2023 IEEE Global Engineering Education Conference (EDUCON)*, pp, 1–6. doi: 10.1109/EDUCON54358.2023.10125267.

YoYo Games Ltd (2023), gamemaker.io/en, v2023.8.1.102.

Zheng, Lanqin, Bhagat, Kaushal Kumar, Zhen, Yuanyi, and Zhang, Xuan (2020). "The Effectiveness of the Flipped Classroom on Students' Learning Achievement and Learning Motivation: A Meta-Analysis", *Journal of Educational Technology & Society,* 23(1), pp. 1–15.