iReflect: Enhancing Reflective Learning with LLMs: A Study on Automated Feedback in Project Based Courses

Bhojan Anand[a] and Quek Sze Long[b]
School of Computing, National University of Singapore, Singapore

Keywords: Pedagogy, Project, Project Based Learning, Experiential Learning, Reflective Learning, Automated Reflection Feedback, Large Language Model, ChatGPT, Deepseek.

Abstract: Reflective learning in education offers various benefits, including a deeper understanding of concepts, increased self-awareness, and higher-quality project work. However, integrating reflective learning into the syllabus presents challenges, such as the difficulty of grading and the manual effort required to provide individualised feedback. In this paper, we explore the use of Large Language Models (LLMs) to automate formative feedback on student reflections. Our study is conducted in the CS4350 Game Development Project course, where students work in teams to develop a game through multiple milestone assessments over the semester. As part of the reflective learning process, students write reflections at the end of each milestone to prepare for the next. Students are given the option to use our automated feedback tool to improve their submissions. These reflections are graded by Teaching Assistants (TAs). We analyse the impact of the tool by comparing students' initial reflection drafts with their final submissions and surveying them on their experience with automated feedback. In addition, we assess students' perceptions of the usefulness of reflective writing in the game development process. Our findings indicate that students who revised their reflections after using the tool showed an improvement in their overall reflection scores, suggesting that automated feedback improves reflection quality. Furthermore, most of the students reported that reflective writing improved their learning experience, citing benefits such as increased self-awareness, better project and time management, and enhanced technical skills.

1 INTRODUCTION

Reflective learning has received greater attention since the 1980s for its potential role in education as a form of self-directed learning. Schön (1984) had demonstrated how reflective learning is rooted within diverse professional contexts, from creative disciplines to science-based fields. Kolb (1984) also identified critical reflection as a core element in experiential learning. In particular, Bhojan and Hu (2024), after integrating reflective writing into project-based game development courses, have found that the quality of student reflections has a positive correlation with the quality of the final submitted project.

While there has been much interest in integrating reflective learning into higher education, there are a number of challenges that hinder its application (Chan and Lee, 2021). One difficulty is the task of objectively assessing student reflections; educators are often not specifically trained in grading reflective writing while also being susceptible to bias. Another difficulty is the additional burden on the educator with the time and effort required for manually grading each and every student's reflections and providing individualized feedback.

With Bhojan and Hu (2024)'s work as the primary motivation, this study aims to develop an automated feedback feature using generative AI and Large Language Models (LLMs) that students can utilise to improve the quality of their reflections, thereby improving the quality of their work. By developing an automated system, each reflection can be assessed consistently without personal bias and timely feedback can be given to students at their own convenience. LLMs also have the added advantage of being able to generate feedback that can be tailored to each student's unique experiences. So far, only two automated feedback systems have been developed specifically for reflective writing, both of which use rule-based AI (Knight et al., 2020; Solopova et al., 2023). However,

[a] https://orcid.org/0000-0001-8105-1739
[b] https://orcid.org/0009-0009-5109-9729

Anand, B. and Long, Q. S.
iReflect: Enhancing Reflective Learning with LLMs: A Study on Automated Feedback in Project Based Courses.
DOI: 10.5220/0013435800003932

In *Proceedings of the 17th International Conference on Computer Supported Education (CSEDU 2025) - Volume 2*, pages 395-403
ISBN: 978-989-758-746-7; ISSN: 2184-5026

no studies have yet explored the use of generative AI or large language models (LLMs) in this context.

The automated feedback system is implemented as a feature in iReflect, a web application tool developed in our university that can facilitate critical peer review, discussions over peer reviews and individual reflections over multiple milestones (Tan, 2022). To study the effectiveness of our automated feedback tool in enhancing reflective learning, we conducted a study in the course CS4350: Game Development Project. CS4350 is a project-based game development course with high demand on creative and technical skills, which has integrated reflective learning as part of its coursework after the study by Bhojan and Hu (2024).

The current study is guided by the following two questions:

1. Does our LLM-based automated feedback tool improve the quality of student reflections?
2. Do the students perceive benefits from accomplishing the reflective tasks? If so, what kind of benefits do they perceive?

2 LITERATURE REVIEW

Reflective learning was first coined by Dewey (1933), who argued that reflection is a necessary process to draw out meaning from experience and to use that meaning in future experiences. Following that, Schön (1984) notably identified two distinct types of reflection: "reflection-in-action", where learners reflect and adjust during an ongoing process, and "reflection-on-action", where learners reflect and analyze after the end of a process. He argues that both reflections are essential to practitioner's practice. Meanwhile, Kolb (1984) had proposed a model of experiential learning and had noted that experience alone does not transform into knowledge, critical reflection is necessary to bring about learning from experience. Subsequently, with the accelerating pace of technological changes in the world, there has been greater interest in reflective learning as a pedagogical approach that builds up student capacities for lifelong learning (Bourner, 2003).

While the benefits of reflective writing have been studied and documented, there have been a number of challenges in integrating it into curriculum. Chan and Lee (2021) reviews many of these challenges found in literature and notes that there are multiple levels of challenges, from the student learning level, to the teacher pedagogical level, institutional level and finally the sociocultural level. On the teacher pedagogical level, one challenging area is the assessment of reflections, where a number of studies found that teachers faced difficulties setting standards to grade reflections and assess them objectively. Bourner (2003) also notes that student reflections involve personal, emergent learning which is hard to assess with a predetermined criteria.

To grade objectively, an objective assessment criteria would be necessary. Currently, there is no singular accepted model for reflective learning as a basis for reflection assessment. A number of different reflections models have been proposed and each has been widely adapted for use, such as Gibbs' Reflective Cycle (Gibbs and Unit, 1988) and Rolfe et al.'s Reflective Model (Rolfe et al., 2001).

One proposed rubric is by Tsingos et al. (2015), who paired the different stages of reflections suggested by Boud et al. (1985) and the different levels of depth of critical reflection as proposed by Mezirow (1991) to make a new matrix rubric for reflective writing in the context of pharmaceutical education. Building upon Tsingos et al.'s work, Bhojan and Hu (2024) then proposed a simplified rubric that omits two of the more complex stages of reflection to improve consistency of grading between human graders while maintaining the reflective learning outcomes, which was used in the context of creative media and game development courses.

Automated feedback and scoring is an area that is still being actively studied, particularly in the area of essay writing. However, there are few studies in the specific context of reflective writing. There has only been two published automated feedback systems tailored for reflective writing, AcaWriter (Knight et al., 2020) and PapagAI (Solopova et al., 2023). AcaWriter is a learning analytics tool developed by Knight et al. to provide feedback on academic writing, including reflective writing. It was developed with the text analysis pipeline by Gibson et al. (2017) and uses a rule-based AI framework to identify the presence of certain literary features that are hallmarks of reflective writing.

Meanwhile, PapagAI is an open source automated feedback tool system developed by Solopova et al. based on didactic theory, implemented as a system of multiple machine learning model modules, each fine-tuned to detect different elements of reflective writing before coming up with an overall feedback regarding lacking elements. These elements include the detection of emotions, identifying which phases of Gibbs' Reflective Cycle (Gibbs and Unit, 1988) are present, and the level of reflections according to the Fleck and Fitzpatrick Scheme (Fleck and Fitzpatrick, 2010).

There are no studies yet on utilizing generative AI and LLMs for automated feedback for reflections.

Some concerns pointed out by Solopova et al. (2023) regarding the use of LLMs is the lack of transparency and control over the output and hallucinations, which was why a rule-based AI using traditional machine learning models was preferred. Still, they note that LLMs do hold great promise and have advantages such as greater speed over a full system of language models.

In the relatively new area of prompting strategies, there has been a great influx of studies in recent years. While there are no studies on prompting in the specific context of reflective writing feedback, there are many studies done on feedback in the context of essay writing or English as a Foreign Language (EFL) learning (Stahl et al., 2024; Yuan et al., 2024; Han et al., 2024).

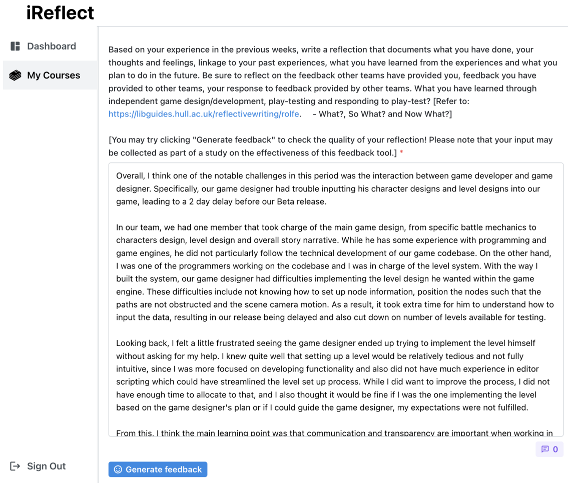

Figure 1: iReflect System: The interface for students to enter their reflection writing with the option to generate feedback.

3 iReflect FRAMEWORK

iReflect (https://ireflect.comp.nus.edu.sg) is an in-house web application tool developed by our university's students that helps educators facilitate critical peer review, discussions over peer reviews and individual reflections over multiple milestones (Tan, 2022). One of our key objectives is to be able to facilitate reflective learning for the student. To this end, iReflect provides an automated reflection feedback feature that can generate timely feedback for a student's reflection at the student's own convenience. Before this study, the automated feedback generation feature was based on AcaWriter Knight et al. (2020). AcaWriter uses traditional machine learning models trained to detect the presence of literary features deemed important for reflective learning. In this study, a new feedback system based on LLM prompting was developed and integrated into iReflect and replaces the use of AcaWriter. A prompt engineering approach was selected over a data-driven approach due to the limited quantity and quality of data available for qualitative feedback for student reflections.

The new feedback system utilises OpenAI's GPT-4o model, which is considered to be at the forefront of developed LLMs. At the time of this paper, the latest GPT-4o model we adopted is gpt-4o-2024-08-06.

In engineering the prompt, we refer to findings on prompting strategies on generating feedback from Stahl et al. (2024), Yuan et al. (2024) and Han et al. (2024).

Stahl et al. (2024) explored different prompting strategies in the context of automated essay writing, which shares many commonalities with reflective writing. They explored the use of personas, various instruction patterns (scoring, feedback, Chain-of-Thought and combinations of the three) and in-context learning (adding examples to the prompt). For the personas, it is noted that the Teaching Assistant persona and Educational Researcher persona performed better than no persona and Creative Writing persona. Among instruction patterns, Feedback, Feedback+Scoring and Feedback with Chain-of-Thought+Scoring produced the most helpful responses in descending order.

Yuan et al. (2024) had found that providing specific guidelines and criteria in generating feedback for paper introduction writing task provides more constructive and helpful responses, with one of the models tested being GPT4. The responses were evaluated by feeding prompts to Claude2, and were also validated by having two of their experienced NLP (Natural Language Processing) researchers grade a subset of samples and comparing the accuracies.

Han et al. (2024) had explored the use of LLM-as-a-tutor in the context of EFL (English as a Foreign Language) learning. They introduce educational metrics specifically designed to assess student-LLM interactions within the context of EFL writing education and use them as a basis for assessing and comparing the feedback given by standard prompting and score-based prompting. As a result, it is found that score-based prompting generates more negative, straightforward, direct, and extensive feedback than standard prompting, which are preferred attributes for student learners. This is also corroborated by majority of teacher annotators prefer the former over the latter.

In summary, we utilise score-based feedback prompting to generate more negative, straightforward and direct feedback (Han et al., 2024) and explicitly describe the full criteria for reflection assessment to improve the constructiveness of feedback (Yuan et al., 2024). Additionally, we utilise the strategies of adopting a persona of an educator and requesting for feedback and scoring with chain of thought structure (Stahl et al., 2024).

Bhojan and Hu (2024) 's reflection assessment rubric (see Table 1) was used as the basis for the LLM to evaluate and generate feedback for student reflections. The usage of the rubric in tandem with the reflective writing tasks has been found to improve the reflection quality and project quality of students in creative media and game development courses (Bhojan and Hu, 2024).

After writing their reflection, the students have the option to generate feedback using the automated feedback system. The generated feedback is displayed below their reflection as shown in Figure 2.

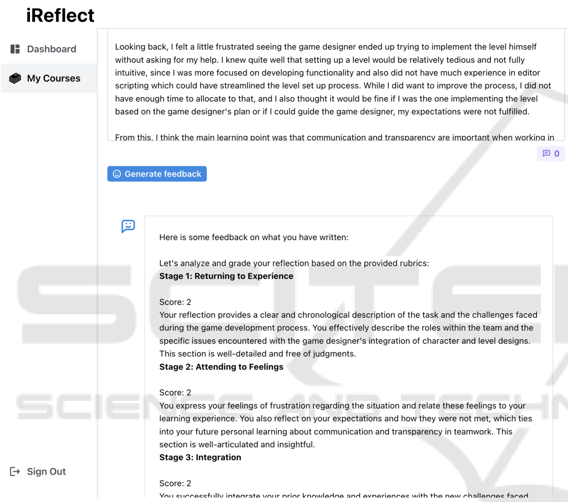

Figure 2: iReflect System: The display of feedback generated for the student's reflection.

4 STUDY AND ANALYSIS

4.1 Course Selected for the Study

The study was conducted in the game development course: CS4350 Game Development Project in our university. The course is project-based, aiming to provide hands-on practical experience in game development by having students form teams and work on developing a complete game from start to end throughout the semester. The final game product accounts for 45% of the total grades in the course. Lectures on game development practices were given in the first half of the course, while the rest of the time was dedicated to allowing students work on their project. Apart from a selected theme (for example, the theme for this iteration was "Serious Games"), the project is open-ended and the students have the freedom to decide the content.

Throughout the course, to track progress and provide formative feedback on the work so far, the students have 5 milestone assessments in total: *Concept Phase*, *Prototype Phase*, *Alpha Phase*, *Beta Phase* and *Gold Phase*. For each milestone assessment, teams are to prepare a presentation for the rest of the class, showcasing their current progress. For milestones from Alpha Phase onward, teams are also expected to prepare a playable version of their game for other teams to play-test. After the presentations and play-testing the games , teams will then critically peer-review each other. Teams can respond to these peer reviews and participate in discussion before deciding whether to accept or reject the other team's suggestions. Finally, every student is tasked to write an individual reflection on their overall experience working towards the latest milestone. The peer review process and the reflective writing tasks are hosted on our web application tool iReflect.

For the reflection task, the students are tasked to write a single short reflection essay based on a given prompt. One example of the prompts used:

> "Based on your experience in the previous weeks, write a reflection that documents what you have done, your thoughts and feelings, linkage to your past experiences, what you have learned from the experiences and what you plan to do in the future. Be sure to reflect on the feedback other teams have provided you, the feedback you have provided to other teams and your response to feedback provided by other teams. What have you learned through independent game design/development, play-testing, and responding to play-test?"

The submitted reflections are graded by the two TAs for CS4350 using the rubric proposed by Bhojan et al. (2024) shown in Table 1.

4.2 Study Methodology

A total of 26 students participated in the study. They were first introduced to the automated reflection feedback tool during the **Prototype Phase** and given the option to use it. Students were informed that using the tool was voluntary, would not affect their grades, and that their responses would be collected for research if they chose to participate. Reflection responses were then gathered during the **Alpha Phase** and **Beta Phase**, with all submissions collected after the respective deadlines. For students who used the feedback tool, their initial draft—the version submitted for feedback—was also collected for analysis.

After the collection of initial and final reflection

Table 1: Six Stage Rubrics for Reflection Statement Assessment in Project Based Courses (Bhojan and Hu, 2024).

Rubric	Nonreflector (0 Marks)	Reflector (1 Mark)	Critical Reflector (2 Marks)
Stage 1: Returning to Experience	Statement does not provide a clear description of the task itself.	Statement provides a description of the task.	Statement provides description of the task chronologically and is clear of any judgments.
Stage 2: Attending to Feelings	Statement provides little of no evidence of personal feelings, thoughts.	Statement conveys some personal feelings and thoughts of the clinical experience but does not relate to personal learning.	Statement conveys personal feelings, thoughts (positive or negative) of the experience and relates to future personal learning.
Stage 3: Integration	Statement shows no evidence of integration of prior knowledge, feelings, or attitudes with new knowledge, feelings or attitudes, thus not arriving at new perspectives.	Statement provides some evidence of integration of prior knowledge, feelings, or attitudes with new knowledge, feelings or attitudes, thus arriving at a new perspective.	Statement clearly provides evidence of integration of prior knowledge, feelings, or attitudes with new knowledge, feelings or attitudes, thus arriving at new perspectives.
Stage 4: Appropriation	Statement does not indicate appropriation of knowledge.	Statement shows appropriation of knowledge and makes inferences relating to prior inferences and prior experience.	Statement clearly shows evidence that inferences have been made using their own prior knowledge and previous experience throughout the task.
Stage 5: Outcomes of Reflection	Statement shows little or no reflection on own work, does not show how to improve knowledge or behavior, and does not provide any examples for future improvement.	Statement shows some evidence of reflecting on own work, shows evidence to apply new knowledge with relevance to future practice for improvement of future practice. Provides examples of possible new actions that can be implemented most of the time.	Statement clearly shows evidence of reflection and clearly states: (1) a change in behavior or development of new perspectives as a result of the task; (2) ability to reflect on own task, apply new knowledge, feelings, thoughts, opinions to enhance new future experiences; and (3) examples.
Stage 6:[1]Readability and Accuracy	Difficult to understand, includes errors in spelling, grammar, documentation, and/or inaccurate key details.	Accurate, understandable text, includes all key details.	Clear, engaging, accurate and comprehensive text.

[1] *Readability and Accuracy* - To what extent does this reflection convey the effect of the learning event?

responses, the responses were coded and compiled together in a randomized order before being graded by the TAs of CS4350. This is to allow the graders to grade both initial and final responses equally without bias.

At the end of each reflection task in Alpha and Beta Phase, the students were given survey questions to gather their self-perceived effects of the self-reflection task on their learning experience. The following survey questions were asked:

1. To what extent do you agree that the learning reflection in the Prototype phase has helped you in completing tasks in the Alpha phase?
2. If you agreed with the previous statement, what are some areas that the learning reflection has helped in this project? (E.g. greater self-awareness, deeper understanding of concepts, new perspectives, etc.) If you disagreed with the previous statement, what are some of the reasons you found it unhelpful?
3. To what extent do you agree that engaging in learning reflection improved your overall learning experience during this project so far?
4. If you agreed with the previous statement, what are some areas that the learning reflection has helped you overall, both within and outside of the course? (E.g. greater self-awareness, deeper understanding of concepts, new perspectives, etc.) If you disagreed with the previous statement, what are some of the reasons you found it unhelpful?

For Questions 1 and 3, the students were given a 7-point Likert scale to indicate their level of agreement. For Questions 2 and 4, the response was open-ended to allow students to freely express their perceived benefits or detriments.

Lastly, at the end of the final milestone Gold Phase, students were given the following survey questions to gather their opinions on the automated feedback feature:

1. Rate your agreement with the following statement: "The automated feedback probed me to think and reflect more deeply."
2. Which of the following aspect(s) of the automated feedback do you find helpful?

   [Available Options: Concrete suggestions, Specific comments, Good balance of both positive and negative comments, Stimulating questions, Others]

3. What are some problems / areas for improvement that you identify in the automated feedback?
4. If you did not use the automated feedback feature, what are the reasons for not using it?

4.3 Study Results

A total of 26 students signed up for CS4350. In total, 20 and 22 submissions were collected for the reflection task for Alpha Phase and Beta Phase respectively, giving a respective submission rate of 76.9% and 84.6%.

Out of the submissions, students were differentiated into whether they used the automated feedback feature or not. Additionally, students that used the feature were further differentiated by whether they made changes between the initial response and final response after using the feedback feature. The distribution of students across the three groups is detailed in Table 2. Subsequently, the mean score of students in each group was calculated for each scoring category and presented in Tables 3 and 4.

Table 2: Student Usage of Automatic Feedback Feature across Alpha and Beta Phase.

Usage Type	Alpha Phase	Beta Phase
Used Feedback Feature	16	16
Made Changes	9	9
No Change	7	7
Did Not Use Feedback	4	6
Total Submissions	20	22

The results of the survey on the students' opinions of the automated feedback feature are collated in Figures 3 and 4. Figure 3 shows the distribution of responses on a 7-point Likert scale regarding agreement to the statement "The automated feedback probed me to think and reflect more deeply." Figure 4 lists aspects of constructive feedback with the number of students that found that aspect present in the automated feedback and felt that it was useful.

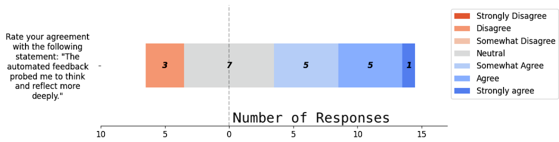

Figure 3: Gold Phase Survey Results on agreement with the statement: "The automated feedback probed me to think and reflect more deeply.".

Finally, the results of the survey on students' perceived feedback from the reflective writing task are detailed in Figures 5 and 6 for Alpha Phase and Beta Phase respectively.

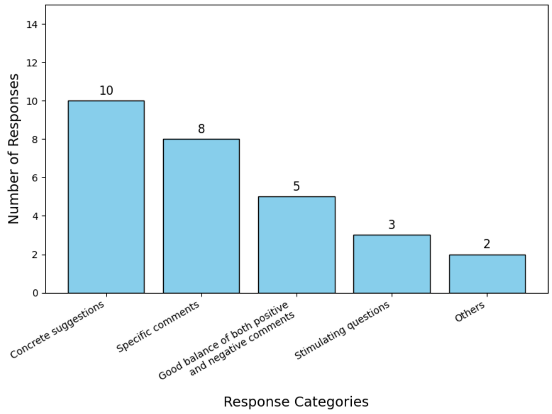

Figure 4: Gold Phase Survey Results on aspect(s) of the automated feedback that were found useful.

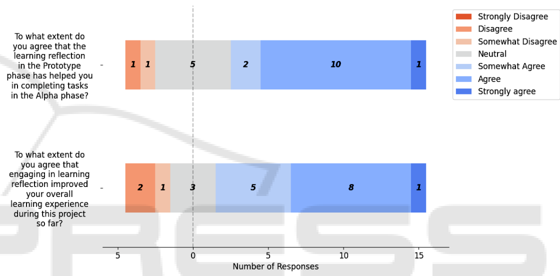

Figure 5: Alpha Phase Survey Results.

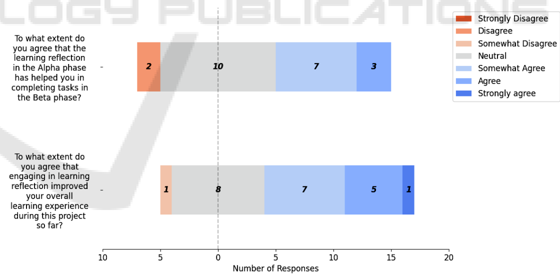

Figure 6: Beta Phase Survey Results.

4.4 Study Results Analysis

4.4.1 Reflection Scores

From the final reflection score statistics in Tables 3 and 4, students that used the feedback feature, whether they made changes after feedback or not, achieved higher scores than students that did not use the feedback feature at all. It is also seen that students generally scored the lowest in the reflection stage Stage 4: Appropriation, regardless of feature usage.

Table 3: Statistics of Final Reflection Scores by Feature Usage in Alpha Phase.

Scoring Category	**Mean Users (Made Changes)**	**Mean Users (No Changes)**	**Mean Non-Users**	**Overall Mean**
Stage 1	1.78 ± 0.36	1.83 ± 0.26	1.88 ± 0.25	1.82 ± 0.30
Stage 2	1.72 ± 0.44	1.58 ± 0.58	1.38 ± 0.25	1.61 ± 0.46
Stage 3	1.44 ± 0.53	0.92 ± 0.49	0.88 ± 0.75	1.16 ± 0.60
Stage 4	1.17 ± 0.66	0.75 ± 0.61	0.75 ± 0.65	0.95 ± 0.64
Stage 5	1.56 ± 0.53	1.50 ± 0.45	1.38 ± 0.25	1.50 ± 0.44
Readability and Accuracy	1.89 ± 0.22	1.83 ± 0.26	1.75 ± 0.50	1.84 ± 0.29
Total Score	9.56 ± 1.47	8.42 ± 1.99	8.00 ± 1.83	8.87 ± 1.75

Table 4: Statistics of Final Reflection Scores by Feature Usage in Beta Phase.

Scoring Category	**Mean Users (Made Changes)**	**Mean Users (No Changes)**	**Mean Non-Users**	**Overall Mean**
Stage 1	1.83 ± 0.35	1.93 ± 0.19	1.75 ± 0.42	1.84 ± 0.32
Stage 2	1.72 ± 0.44	1.43 ± 0.45	1.50 ± 0.45	1.57 ± 0.44
Stage 3	1.72 ± 0.51	1.71 ± 0.27	1.42 ± 0.74	1.64 ± 0.52
Stage 4	1.56 ± 0.53	1.43 ± 0.53	1.25 ± 0.61	1.43 ± 0.54
Stage 5	1.72 ± 0.36	1.79 ± 0.27	1.33 ± 0.82	1.64 ± 0.52
Readability and Accuracy	1.89 ± 0.22	1.93 ± 0.19	1.67 ± 0.52	1.84 ± 0.32
Total	10.44 ± 1.49	10.21 ± 1.32	8.92 ± 2.25	9.95 ± 1.72

Table 5: Paired t-test Results for Initial and Final Reflection Scores for Alpha Phase.

Scoring Category	**Mean Initial Score**	**Mean Final Score**	**Change in Score**	**t-value**	**p-value**
Stage 1	1.44	1.78	+0.34	2.828	0.022*
Stage 2	1.17	1.72	+0.55	2.626	0.030*
Stage 3	0.50	1.44	+0.94	4.857	0.001**
Stage 4	0.44	1.17	+0.73	3.043	0.016*
Stage 5	0.78	1.56	+0.78	3.092	0.015*
Readability and Accuracy	2.00	1.89	-0.11	-1.512	0.169
Total Score	6.33	9.56	+3.23	4.685	0.002**

Note: * indicates p < 0.05, ** indicates p < 0.01.

Table 6: Paired t-test Results for Initial and Final Reflection Scores for Beta Phase.

Scoring Category	**Mean Initial Score**	**Mean Final Score**	**Change in Score**	**t-value**	**p-value**
Stage 1	1.89	1.83	-0.056	-1.000	0.35
Stage 2	1.33	1.72	+0.39	2.135	0.065
Stage 3	1.33	1.72	+0.39	2.135	0.065
Stage 4	1.17	1.56	+0.39	2.135	0.065
Stage 5	1.39	1.72	+0.33	2.828	0.022*
Readability and Accuracy	1.94	1.89	-0.056	-0.555	0.59
Total Score	9.06	10.44	+1.39	2.786	0.024*

Note: * indicates p < 0.05.

A paired t-test was conducted to examine the differences between initial and final reflection scores across the different stages described by the assessment rubric. Table 6 shows the mean initial scores, the mean final scores, t-values and p-values for each category.

In the Alpha Phase, it is seen that across the 5 stages of reflection and in total, there has been an increase in score from the initial reflection to the final reflection ($p < 0.05$). In particular, the students generally had scores below 1 for Stage 3, 4 and 5, which is below the standard for a basic reflector in the initial reflection. However, after making changes with the help of the feedback feature, they were able to improve their scores to well above a score of 1.

In the Beta Phase, students generally had a higher score across all stages of reflection in their initial reflection as compared to in Alpha Phase. Subsequently, it is seen that there is a smaller improvement in scores across all stages of reflection, with Stage 1 even having a slight reduction in score. In Total, there is still an improvement in score ($p < 0.05$).

The above results suggest that overall, the automated feedback feature has improved the quality of student reflections. While significant improvements in scores were seen in Alpha Phase, these improvements were much smaller in Beta Phase, which suggests that the students did not have a clear understanding of the different reflection stages when first doing reflections in Alpha Phase until the first round of automated feedback.

On the other hand, the readability and accuracy category does not show any conclusive changes between initial and final ($p > 0.05$) in both Alpha and Beta Phases. This suggests that our feedback feature does not provide helpful feedback in terms of improving the overall coherence and flow of the students' reflections.

4.4.2 Survey Results

From the Gold Phase survey results in Figure 3, it is seen that slightly more than half of the students agree that overall, the automated feedback directed them to think and reflect more deeply, which suggests that the automated feedback has been relatively successful in encouraging deeper reflective learning. From Figure 4, some of the strengths of the automated feedback are the concrete suggestions (47.6% of responses, n = 10) and specific comments (38.1% of responses, n = 8) provided in the feedback.

However, it is noted that there is still a sizable number of students who do not find the feedback useful in encouraging deeper reflection, replying with neutral or even disagreement to the first question. Among these responses, there are various reasons provided, the most common reason was that the feature which used a rubric as its basis of feedback felt too "rigid" and "formulaic". They felt that the feedback system was just "ticking off criteria" without encouraging "deeper exploration or improvement in the quality and depth of the content", and that these sections may not always be relevant to all reflections. This relates back to Bourner (2003)'s theory that student reflections involve personal, emergent learning which is hard to assess with a predetermined criteria. Another point of improvement that students mentioned is that inconsistency of the feedback, which may still provide different scores and replies given the same or just slightly different input.

From the survey results in Figures 5 and 6, it is seen that majority of the students agree that the self-reflection tasks had improved their learning experience both for the following milestone and overall for the project so far.

Some common areas that they felt the reflection has helped include:

1. **Self-awareness and Identification of Strengths and Weaknesses**
   Students gained more self-awareness about their strengths, weaknesses, and learning habits, helping them adjust their approach in future work.
2. **Improved Collaboration and Team Dynamics**
   Reflection enhanced their understanding of individual contributions and team dynamics, leading to better collaboration.
3. **Better Project or Time Management**
   Reflecting on past experiences improved project management and task prioritisation.
4. **Improvement in Technical Skills**
   Students refined their technical approaches, adapted to new tools, and leveraged prior experience to solve problems.

Meanwhile, among disagreeing responses, the most common complaint is that the reflection tasks were too frequent within the project time frame. While they generally find the self-reflection task is useful, they feel that 2-3 weeks between each reflective task is too short of a time to have enough meaningful experience to reflect, reducing the effectiveness of the reflection and making task itself more tedious. This seems to be supported by how the number of responses agreeing that the reflective task improved their learning experience decreased from Alpha Phase to Beta Phase, when the students felt that the time between Alpha Phase and Beta Phase reflections was too short.

From the remaining disagreeing responses, one student mentions that the self-reflection task did not bring additional benefit since they already reflect in their own time.

From these responses, we note that when integrating reflective learning into coursework, reflection tasks should be spaced adequately to allow students to gather meaningful experiences. Too frequent reflections might compromise the effectiveness of the reflection and provide additional workload that detracts students' learning from the rest of the course. Additionally, we note that reflective learning is not always limited to reflective writing in class.

Overall, a majority of the students still do find reflective writing to be helpful, suggesting that it is beneficial overall to include reflective writing in the curriculum for a game development course.

5 LIMITATIONS AND FUTURE WORK

Our implementation is limited by the lack of high-quality reflection feedback data for training and fine-tuning. Enlisting trained human graders to provide gold-standard data would help establish a stronger benchmark for LLM-generated feedback.

Another challenge is the use of OpenAI's GPT model, which, while powerful, operates as a black-box system. Unlike rule-based AI, its feedback generation lacks transparency, making theoretical soundness difficult to verify. As such, it is best suited for formative feedback, complemented by human review.

Lastly, using a third-party LLM raises privacy concerns. To mitigate risks, we restricted reflections to coursework-related content, excluding personal data. However, one student still cited privacy concerns as a reason for avoiding the tool.

6 CONCLUSIONS

By leveraging LLMs, we developed an automated feedback system that provided students with timely, personalized insights on their reflections. Findings from the CS4350 course study show that the system significantly improved reflection quality, with students appreciating its concrete suggestions and specific feedback. While some noted minor drawbacks—such as occasional rigidity and formulaic responses—the overall reception was highly positive, reinforcing its value in the learning process.

Survey results further highlight the benefits of reflective writing, with students reporting increased self-awareness, improved teamwork, better project management, and enhanced technical skills. These findings affirm the value of reflective learning in project based courses. Integrating automated feedback can further enrich student learning, provided that reflective tasks are scheduled thoughtfully to ensure meaningful engagement without adding excessive workload.

REFERENCES

Bhojan, A. and Hu, Y. (2024). Play testing and reflective learning ai tool for creative media courses. In *Proceedings of the 16th CSEDU - Volume 1*, pages 146–158. INSTICC, SciTePress.

Boud, D., Keogh, R., and Walker, D. (1985). Promoting reflection in learning: A model. In *Reflection: Turning Reflection into Learning*. Routledge, London.

Bourner, T. (2003). Assessing reflective learning. *Education + Training*, 45(5):267–272.

Chan, C. and Lee, K. (2021). Reflection literacy: A multilevel perspective on the challenges of using reflections in higher education through a comprehensive literature review. *Educational Research Review*, 32:100376.

Dewey, J. (1933). *How We Think: A Restatement of the Relation of Reflective Thinking to the Educative Process*. D.C. Heath and Company, New York.

Fleck, R. and Fitzpatrick, G. (2010). Reflecting on reflection: framing a design landscape. In *Proceedings of the 22nd Conference of the Computer-Human Interaction Special Interest Group of Australia on Computer-Human Interaction*, OZCHI '10, page 216–223, New York, NY, USA. Association for Computing Machinery.

Gibbs, G. and Unit, G. B. F. E. (1988). *Learning by Doing: A Guide to Teaching and Learning Methods*. FEU.

Gibson, A., Aitken, A., and et al. (2017). Reflective writing analytics for actionable feedback. In *Proceedings of the Seventh International Learning Analytics & Knowledge Conference*, LAK '17, page 153–162, New York, NY, USA. Association for Computing Machinery.

Han, J., Yoo, H., Myung, J., and et al. (2024). Llm-as-a-tutor in efl writing education: Focusing on evaluation of student-llm interaction. KAIST, South Korea.

Knight, S., Shibani, and et al. (2020). Acawriter: A learning analytics tool for formative feedback on academic writing. *Journal of Writing Research*, 12:141–186.

Kolb, D. A. (1984). *Experiential Learning: Experience as the Source of Learning and Development*. FT Press.

Mezirow, J. (1991). *Transformative Dimensions of Adult Learning*. Jossey-Bass, San Francisco.

Rolfe, G., Freshwater, D., and Jasper, M. (2001). *Critical Reflection for Nursing and the Helping Professions: A User's Guide*. Palgrave MacMillan.

Schön, D. A. (1984). *The Reflective Practitioner: How Professionals Think in Action*, volume 5126. Basic Books.

Solopova, V., Rostom, E., and et al. (2023). Papagai: Automated feedback for reflective essays. In Seipel, D. and Steen, A., editors, *KI 2023: Advances in Artificial Intelligence*, pages 198–206, Cham. Springer Nature Switzerland.

Stahl, M., Biermann, L., Nehring, A., and Wachsmuth, H. (2024). Exploring llm prompting strategies for joint essay scoring and feedback generation. Leibniz University Hannover.

Tan, K. Q. J. (2022). Playtesting and reflective learning tool for creative media courses. B. Comp. Dissertation, Project Number H1352060, 2021/2022.

Tsingos, C., Bosnic-Anticevich, S., Lonie, J. M., and Smith, L. (2015). A model for assessing reflective practices in pharmacy education. *American Journal of Pharmaceutical Education*, 79:124.

Yuan, W., Liu, P., and Gallé, M. (2024). Llmcrit: Teaching large language models to use criteria.