Play Testing and Reflective Learning AI Tool for Creative Media Courses

Bhojan Anand[a] and Hu Yuxin[b]
School of Computing, National University of Singapore

Keywords: Pedagogy, Reflective Learning, Automated Reflection Analysis, Peer Evaluation.

Abstract: Reflective learning has various benefits for students, such as increased motivation, empathy, and academic success. This paper presents a tool for self-regulated reflective learning in creative media courses. We have studied the effect of using the tool in three courses with a strong demand for creativity and technical skills. The courses are CS3247 (Game Development), CS4240 (Interaction Design for Virtual and Augmented Reality), and CS4350 (Game Development Project). Each project team is considered an indie game development studio and adopts best practices from the industry. The projects will go through multiple iterations during the semester, which are graded. Reflective learning is implemented by allowing students to reflect on their past experiences between each iteration. Reflections are assessed by teaching assistants to evaluate the quality of reflections based on standardized rubrics. Two rounds of the study were conducted over 2021/2022 Semester 2 and 2022/2023 Semester 1. The first round was a pilot study, and the feedback from the first round was used to reflect on the experimental methodology and modify it for the second round. One key finding was that free-form reflection writing with guidelines had pushed for better quality reflections than a more specific set of questions. Overall, we have shown positive correlations between the mean student reflection scores in a team and the marks their team earned for submissions of each iteration with moderate inter-rater reliability. In addition, implementing reflective learning in these courses has increased the overall module and teaching feedback scores.

1 INTRODUCTION

Reflective learning is a pedagogical method in which learners actively reflect on their learning experiences to develop new insights. Researchers have discussed the role that reflections play in learning. In Schon (1984)'s work on reflective practice, he encouraged the use of reflection in action to learn by doing. Kolb (1984)'s work on experiential learning suggests that reflection exists as a process to help learners process experiences into abstract concepts. Creative media courses such as game development, VR, and AR experiences development courses in our school teach students design skills and mastery of 3D game development engines such as Unity3D and Unreal Engine. The project-driven nature of the curriculum results in most of the students' learning occurring outside the classroom, where they engage in largely self-directed explorations of these engines. As such, students may learn by consolidating their insights by reflecting on their experiences.

This study aims to develop an intelligent tool to aid students in their reflective learning process that can be integrated into any creative media course without modifying the existing course structure. While there are many previous studies on improving reflection writing skills, there is no sufficient study on incorporating reflection writing as a part of formative assessment components in regular computer science and engineering courses at tertiary level education. This experimental study aims to duplicate the benefits of reflective learning by augmenting it with three courses: CS3247 (Game Development), CS4240 (Interaction Design for Virtual and Augmented Reality), and CS4350 (Game Development Project). As it will add additional workload for students and the teaching team, we have built a software tool to facilitate and automate the entire process and make integrating reflective learning with regular courses feasible. The software also has the intelligence to help students write better reflections. The results are used to verify the feasibility and usefulness of the approach. The approach is applicable to other courses that are project-based and require a high amount of self-directed learning outside the classroom.

[a] https://orcid.org/0000-0001-8105-1739
[b] https://orcid.org/0000-0002-1685-6159

Anand, B. and Yuxin, H.
Play Testing and Reflective Learning AI Tool for Creative Media Courses.
DOI: 10.5220/0012633600003693

In *Proceedings of the 16th International Conference on Computer Supported Education (CSEDU 2024) - Volume 1*, pages 146-158
ISBN: 978-989-758-697-2; ISSN: 2184-5026

The key contributions of the paper are listed below. 1) Incorporated reflective learning practice into creative media courses in our University without modifying the structure of the course. We show that implementing reflective learning within the course activities is feasible and useful in improving students' quality of work. 2) We show that automated feedback on reflection writing improves not only the quality of reflection but also the quality of project work produced by the students. There is a direct correlation between the quality of reflection and project work, as indicated by their grades. 3) Development and analysis of an intelligent tool that automates the entire process with instant feedback on reflection writing.

The rest of the sections are organized as follows. Section 2 discusses related literature, Section 3 presents the research methodology and results for the preliminary study, and Section 4 presents the subsequent study with improvements based on self-reflection on our preliminary study. Section 4 also presents iReflect [1] tool/system which is used in the study. Section 5 discusses additional studies with data from recent semesters and future work. Section 6 concludes with a list of findings and recommendations.

2 LITERATURE REVIEW

Many scholars have theorized about the role of reflection in learning. Dewey (1933) first proposed it with his famous adage, "We do not learn from experience... we learn from reflecting on experience," in his foundational work on teaching and learning. Since then, critical self-reflection has been increasingly recognized as crucial to developing independent learners. Within the experiential learning model by Kolb (1984), reflection is vital in helping people internalize what they have learned and transform their experiences into abstract concepts. In Schon (1984) work on the 'reflective practitioner,' he raised the difference between 'reflection-in-action,' where learners reflect during an action, and 'reflection-on-action,' where learners reflect after an action. Schon claims that both types of reflections are essential in the career development of professionals. In most learning models, reflection is vital to learning and extracting meaning from experiences. These early theories pioneered the research into reflective learning and formed the basis for reflective development frameworks later on.

Schön (1987) shows that in many fields, professional development is substantially improved when people gain the ability to engage in reflective learning by using reflective writing as one approach. Empirical research has also been done to quantify the benefits of reflective learning. Yilmaz and Keser (2016) showed that reflective learning significantly increased students' motivation and academic success for remote learners.

An ongoing area of research is the assessment of reflections. Bourner (2003) highlights the importance of forming rigorous and standardized ways of assessing reflective learning. Indeed, Ramsden (2003) shows that assessment drives students' learning as students view their grades as a critical factor in their level of achievement for their courses.

Researchers have developed different tools and frameworks to help formalize what constitutes reflection. In a meta-study, Tsingos et al. (2015) found that models by Boud et al. (1996) and Mezirow (1991) for assessing levels of reflection were the most extensively used models. When these two models are paired together, they can form a more in-depth and robust education tool.

Sumsion and Fleet (1996) and Wong et al. (1995) note that the same reflective writing can result in different interpreted levels of reflection by other assessors. Indeed, one of the main difficulties in assessing reflections is forming a consistent evaluation of highly variable and subjective writings. The manual and laborious process of assessing reflections adds another layer of difficulty. To address the difficulties faced in assessing reflections, researchers have experimented with using text analysis for education by using both rule-based and machine-learning methods. The approach taken by Ullmann (2019) used sentences as the base unit of measurement for reflective content and evaluated for categories of the reflection detection model. In a later study by Gibson et al. (2017), a concept-matching rhetorical analysis framework was used to automatically detect sentences performing three key reflective rhetorical functions: Context, Challenge, and Change. Kovanović et al. (2018) utilized random-forest classification using linguistic indicators extracted from Linguistic Inquiry and Word Count (Tausczik and Pennebaker (2010)) and Coh-Metrix (Graesser et al. (2004)) for arts students' reflective statements (observation, motive and goal). Only the system by Gibson et al. (2017) provided actionable feedback for students.

[1] https://rubric.comp.nus.edu.sg/ and https://ireflect.comp.nus.edu.sg/

3 PRELIMINARY STUDY AND ANALYSIS

3.1 Media Courses Selected for the Study

We have conducted the study in selected media courses at our university. They are CS3247 Game Development, CS4240 Interaction Design for Virtual and Augmented Reality, and CS4350 Game Development Project. These courses are project-based with at least 50% weightage for final project work. The courses are taught through a blended learning strategy where part of the instructional content is delivered online for learning outside the classroom. To ensure students have mastered the content taught, they are assessed with milestone assessments (Mini-Projects/Assignments). These milestone assessments prepare the students with the necessary skills for the final project. After the submission of each milestone assessment, a team of teaching assistants (TAs) will help to assess and grade the project submissions. At the end of the courses, students will have to complete a final project, which will be the culmination of the skills that they have picked up in the course. The final project is open-ended and consists of a complete 3D game in CS3247 and CS4350 and VR/AR simulation/experience in CS4240.

3.2 Preliminary Study Methodology

During 2021/2022 Semester 2, three reflective learning questionnaires were sent to students taking CS3247 and CS4240. The aggregate information of participating students in these courses is provided in Table 1. Participation in the questionnaires is voluntary, and participants can withdraw anytime. These questionnaires were sent after milestone projects were completed. These reflective questions were tailored to fulfill the rubrics provided by Tsingos et al. (2015) with stages 3 and 5 removed to ease marking and improve inter-rater reliability. The amended rubrics are provided in Table 2. The contents of the questionnaires are given in Table 3. The set of questions provided to the students was engineered to match the different stages. Stage 1 is answered by the technical questions and question 1. Stage 2 is answered by questions 3 and 8. Stage 3 is answered by questions 4 and 5. Stage 4 is answered by questions 2 and 6. Stage 5 is answered by questions 7 and 9. In addition to these reflective questions, additional technical questions that are content specific to the Mini-Projects/Assignments were also asked. These additional questions were gathered from the TAs beforehand and served to refresh the student's memory of the technical details. Some examples of these questions are in Table 4.

In the preliminary study, 55 students from CS3247 and 51 students from CS4240 were enrolled in the study. They were informed that after each of their milestone assessments, a voluntary survey would be made available for them to record their reflections. These reflections are graded by the TAs of CS3247 and CS4240. The TAs are provided with the grading rubric and instructed to evaluate students' achievements in each of the stages across all questions. The students who did not submit reflections for a milestone assessment are given a zero for their reflective score for that milestone assessment. It is a severe penalty, but it helps to analyze the importance of self-reflection. The reflective scores provided by the TAs are analyzed with Fliess' Kappa to check for inter-rater reliability. This will provide a metric to measure the consistency of the reflective scores assigned to students.

There are three Mini-Projects (CS3247) and three Assignments (CS4240) as milestone assessments, which are individual works, and one final project, which is group work. Students' final marks are calculated according to a weighted sum of the milestone assessment marks and the group project marks.

At the end of the course, students' feedback on the reflective learning process was collected. The question given to them was, "How effective do you think these reflective questions are in helping you to learn the contents of the module?". Common themes are extracted from the provided feedback and organized into the categories in Table 6 and Table 7. Each student's feedback is analyzed for the presence of common themes and added to a theme's tally when found.

3.3 Preliminary Study Results

Across all Mini-Projects, the reflection submission rate is 40.6% for CS3247. Across all Assignments, the reflection submission rate is 30.1% for CS4240.

The inter-rater reliability is measured next using Fleiss' kappa. For CS3247, the Fleiss' Kappa is calculated to be 0.17 for reflection ratings across all Mini-Projects. For CS4240, the Fleiss' Kappa is measured to be 0.11 for reflection ratings across all Assignments. Across all reflection ratings, the Fleiss' Kappa is 0.15. The detailed breakdown of Fleiss' Kappa for each Mini-Project or Assignment is given in Table 5.

Feedback on the effectiveness of reflective writing was collected from 28 students from CS3247, and 9 students from CS4240. The common themes were

Table 1: Aggregate data of participating students for CS3247 and CS4240.

	CS3247	CS4240
Gender		
No. of male students	44	36
No. of female students	11	15
Courses Distribution		
School of Computing	43	26
Multi Disciplinary Programme	4	15
Faculty of Arts and Social Sciences	0	1
College of Design and Engineering	1	0
Others	7	9

Table 2: Rubrics used for subsequent study.

Reflective Statement Assessment Rubric	**Nonreflector (0 Marks)**	**Reflector (1 Marks)**	**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 or no evidence of personal feelings, thoughts	Statement conveys some personal feelings and thoughts of the experience but does not relate to personal learning	Statement conveys personal feelings, thoughts (positive and 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 and 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
Readability and accuracy [added only for **subsequent study**]: To what extent does this reflection convey the effect of the learning event?	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.

extracted and the number of students mentioning the themes is tabulated. Positive feedback can be found in Table 6, and negative feedback can be found in Table 7.

3.4 Preliminary Study Analysis

Based on the preliminary data, we can observe that the submission rates for CS3247 and CS4240 reflection surveys are both below 50%. Cases of students

Table 3: Set of questions used in the preliminary study.

Question No.	Question
1.	How have I tried to demonstrate my learning and mastery of this content?
2.	If I were to teach what I have learnt to someone else, how would I do it?
3.	Which part of this lesson is more difficult for me? Why is it difficult?
4.	How can the content of this module help me out in other areas? How can what I learn in other modules help me out in this area?
5.	Which part of this course went well? What do I still need to work on?
6.	What would I do differently if I did it again?
7.	Why are the skills I learnt important? How and when can they help me out in the future?
8.	What are some of the interesting things that others have made? What can I learn from them?
9.	How can I improve in the future?

Table 4: Sample content specific questions used in the preliminary study.

Question No.	Question
1.	How do I set up a VR project with Unity?
2.	How do I map my hand movements in-game to the VR controllers?
3.	How do I modify the input manager for the VR controllers and map my actions (e.g. grab/throw) to the respective input?
4.	What are the differences between OnTriggerEnter, OnTriggerExit, and OnTriggerStay methods?

Table 5: Fleiss' Kappa for milestone assessments.

Reflective Score Fleiss' Kappa						
Module	CS3247			CS4240		
Milesstone Assessment	Mini-Project 2	Mini-Project 3	Mini-Project 5	Assignment 1	Assignment 2	Assignment 3
Individual Assessment Kappa	0.13 ($p < 0.05$)	0.19 ($p < 0.05$)	0.20 ($p < 0.05$)	0.12 ($p < 0.05$)	0.07 ($p = 0.14$)	0.11 ($p = 0.08$)
Module Kappa	0.17 ($p < 0.05$)			0.11 ($p < 0.05$)		
Overall Kappa	0.15 ($p < 0.05$)					

Table 6: Number of occurrence of positive feedback themes in student feedback.

Positive feedback themes					
Feedback category	Encourages reflection	Reinforces learning	Encourages improvement	Keeps track of progress	Useful/ Effective
Count	9	7	4	4	10
% of all feedback	24.3%	18.9%	10.8%	10.8%	27.0%
Total	26				

Table 7: Number of occurrences of negative feedback themes in student feedback.

Negative feedback themes						
Feedback category	Repetitive	Too long	Given too early	Tedious	Content can be learnt elsewhere	Not useful
Count	6	2	1	2	1	4
% of all feedback	16.2%	5.4%	2.7%	5.4%	2.7%	10.8%
Total	13					

providing short or non-informative responses are also observed. This is within expectations, as participation in these surveys was not compulsory and had low weightage. Though the students, in general, understood and agreed with the benefits of reflective learning, when it came to practice, it was seen as an additional task and tedious.

The inter-rater reliability of the reflection graders is poor ($\kappa < 0.20$), suggesting that it is difficult for reflection graders to come to a common agreement on how to grade reflections. It is also observed that some students do not seem to understand the purpose of the questions and answer in ways that do not follow the rubric. This might have caused confusion and disagreement among tutors on how well students have met the rubric's criteria. A lack of understanding among students regarding how they are assessed may cause the low-quality response.

From the feedback gathered, a significant fraction of students (approx 20% - 30%) who participated in reflection writing found the process useful, reinforced their learning, and encouraged reflections. However, some students (approx 16%) found the reflection questions repetitive, tedious, and useless. This may dissuade students from completing their reflections, resulting in lowered submission rates.

The histograms of CS3247's and CS4240's total reflection scores reveal peaks of entries at zero reflection scores, indicative of the high number of students who did not submit the reflective report. Disregarding the entries of zero reflection scores, there is a peak of 5 out of a total of 20 for CS4240 and a peak of 7 out of 20 for CSS3247. Using the Shapiro-Wilk Test, CS3247 shows a normal distribution ($p = 0.09$). In contrast, CS4240 does not show a normal distribution ($p < 0.05$) with a general skew to the lower end of the score (skewness of 1.53), indicative of a high degree of asymmetry. This indicates that the reflection submissions for CS3247 and CS4240 are low quality.

We observed no significant correlation between reflection scores and the student's marks. It is hypothesized that the lack of significant correlation is due to the following factors:

1. The tedious nature of filling out survey forms causes students to refrain from participating in reflections or providing ineffectual reflections.
2. Lack of guidance and feedback on how to write good reflections, resulting in ineffectual reflections.
3. Low-weightage and non-compulsory nature of reflection resulting in students not putting effort into reflection writing.

3.4.1 Comparisons with Similar Studies

Epp et al. (2019)'s study investigated the effects of reflective writing practices and peer feedback on students' reflective writing and writing quality scores in a computer science education context. They found no increase in students' reflective writing quality through peer feedback. This finding was replicated in our preliminary study, where the quality of student reflections was poor, and no correlation between reflection scores and marks achieved was found. We believe that the cause of this lack of improvement was students not knowing how to write effective reflections due to the lack of rubrics provided to them. According to Cheng and Chan (2019), students with access to grading rubrics for reflections generally showed much better reflective writing skills than those without access to reflective writing rubrics.

4 SUBSEQUENT STUDY AND ANALYSIS

4.1 Subsequent Study Methodology

Table 8: Aggregate data of participating students for CS4350.

	CS4350
Gender	
No. of male students	33
No. of female students	7
Courses Distribution	
School of Computing	30
Multi Disciplinary Programme	1
Faculty of Arts and Social Sciences	3
Others	6

The following semester, a similar study was conducted on CS4350 Game Development. The aggregate information of participating students in CS4350 is provided in Table 8. This module aims at teaching students how to develop a game following the main stages of the game development process. The students in teams will gradually develop a game through multiple phases: Concept Phase, Prototype Phase, Alpha Phase, Beta Phase, and Gold Phase. The milestone assessments were conducted at the Prototype, Alpha, Beta, and Gold Phases. Each milestone assessment is critically play-tested/peer-reviewed by other teams, and the teams will respond to the peer-reviews. The response will trigger a discussion between the two teams, resulting in either accepting or rejecting the suggestion. Similar to the preliminary study, three student TAs with backgrounds in game development

will grade the reflective writing of all students at the end of each phase.

The preliminary study provided valuable information that guided the design of our subsequent study. The following amendments are made to the methodology:

Reflections were no longer organized in the form of multiple questions but in the form of a single short essay. This gave students the flexibility to write about what they deemed important, hoping to make the process less tedious. The prompt provided to the students is as follows: "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?"

Reflective writing is now a graded component of the coursework. This ensures that students are motivated to complete the reflection and have higher submission rates (Bourner, 2003). Students must submit a compulsory reflection at the end of the Prototype, Alpha, Beta, and Gold Phases. Three TAs will then mark the reflections based on the provided rubrics.

Students were given access to the iReflect system (refer Section 4.2), which provides automated feedback for their reflections. Acawriter is a learning analytics tool developed by Knight et al. (2020) to provide feedback on academic writing. It is developed with the text analysis pipeline by Gibson et al. (2017). Based on Acawriter, a similar system was implemented to provide students with instantaneous feedback for their reflections. This system was developed using the open-sourced repository of Acawriter and used the same underlying text analysis pipeline for capturing student reflection data.

Students were required to attend a compulsory session at the beginning of the semester on the benefits of writing reflections and how to use automated reflection feedback. This ensures that students understand the rationale of writing reflections, what constitutes reflective writing, and how instructors would grade them. Students were given rubrics for their reflective writing grading to gain awareness of the grading criteria. This followed guidelines of Muncy (2014) by ensuring students know what is expected of them. This intervention was also supported by Cheng and Chan (2019)'s study, which showed that students receiving clear rubrics for assessing reflections displayed better reflection writing skills. To ensure that students submit good writing, an additional section was added to the rubrics to account for the quality of writing. The updated rubrics are shown in Table 2.

Pre-course and post-course surveys were sent to students to measure the observed change in attitude toward reflective writing before and after the course. The students were asked the following questions:

1. Writing reflections enhances my self-directed learning skills.
2. Writing reflections helps me learn from my experiences.
3. Writing reflections is relevant for game development education.
4. Writing reflections helps me plan out what to do in the future.

The participants were provided a 6-point Likert scale to indicate their agreement with the statement. The options provided were Strongly Disagree, Disagree, Slightly Disagree, Slightly Agree, Agree, and Strongly Agree.

At the end of the course, the participants were asked to rate the effectiveness of reflective learning in their game development education and to provide general feedback on the reflective writing process.

4.2 iReflect: An AI Reflective Learning Tool

Based on the lessons from the preliminary study, we have developed a tool (iReflect [2]) which integrates critical peer review, discussions on review and reflection writing with Artificial Intelligence (AI) to provide automated instant feedback. Key features of the system are depicted in Figure 1 and 2.

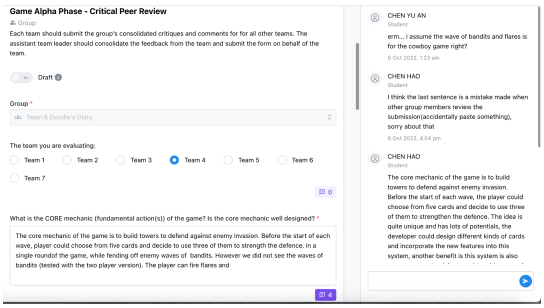


Figure 1: iReflect System: Critical peer review by team 6 to team 4.

[2]https://rubric.comp.nus.edu.sg/ and https://ireflect.comp.nus.edu.sg/

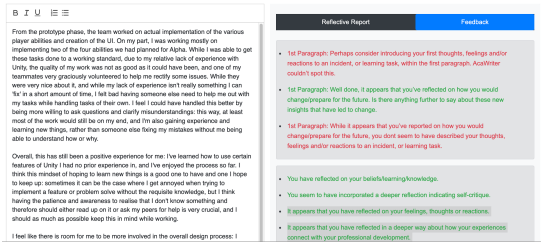


Figure 2: iReflect System: Automated feedback for a reflection statement.

4.3 Subsequent Study Results

A total of 40 students signed up for CS4350. The pre-course survey was sent to students at the start of the semester and garnered 25 responses, giving a response rate of 62.5%. These results of the pre-course survey are found in Figure 3.

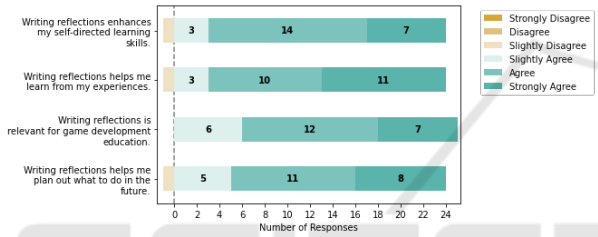


Figure 3: Pre-course survey result.

The post-course survey was sent out at the end of the semester and gathered 24 responses, giving a response rate of 60.0%. Figure 4 shows the post-course survey results.

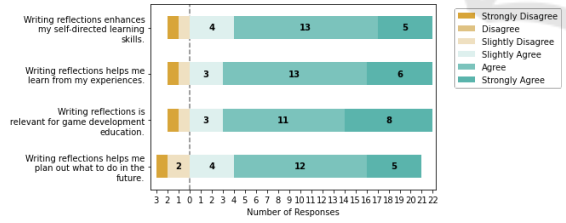


Figure 4: Post-course survey result.

The students were asked how effective reflective writing is in helping them learn the module's contents. Their responses are recorded on a 5-point Likert scale. Their responses are recorded in Figure 5.

The submission rates for the Prototype, Alpha, Beta, and Gold Phases have been collected. The prototype Phase has 34 reflection submissions, resulting in a submission rate of 85%. Alpha Phase has 32 reflection submissions, resulting in a submission rate of 80%. The beta phase has 32 reflection submissions, resulting in a submission rate of 80%. The gold phase has 32 reflection submissions, resulting in a submission rate of 80%.

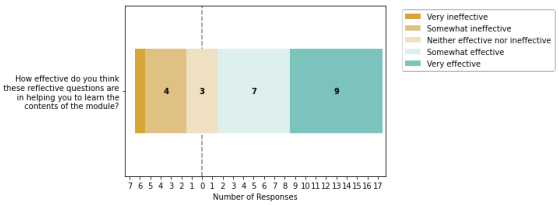


Figure 5: Post-course survey result.

Inter-rater reliability for Prototype, Alpha, Beta, and Gold phase reflections are 0.49, 0.60, 0.55, and 0.56, respectively. Across all phases, the inter-rater reliability is 0.55. This data is also documented in Table 9.

The histogram of reflection submission scores can be found in Figure 6.

For CS4350, only marks for the entire team are available for the Alpha and Beta Phases. Individual marks are only provided for the Gold Phase, where their peer evaluations are also considered. The formula for the final marks for the module is calculated as Alpha Phase marks $\times$ 0.2 + Beta Phase marks $\times$ 0.3 + Gold Phase marks $\times$ 0.5. The marks achieved by each team are shown in Table 10.

The correlation plot between students' reflection scores and marks for the different phases is shown in Figure 7 (low significance $(p > 0.05)$ are masked).

4.4 Subsequent Study Results Analysis

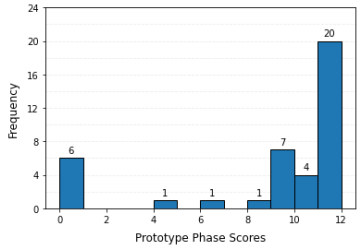
The submission rate and submission quality of student reflections have increased in CS4350. Submission rates have improved to at least 80% across all phases. The increased submission rate is due to multiple factors, including higher weightage and higher emphasis through compulsory training on reflection writing, streamlining the process with the iReflect system that integrates reflection writing with peer-review and discussions process, automated feedback, and the compulsory nature of the reflective writings in CS4350.

From the histograms in Figure 6, the reflection scores do not show a normal distribution $(p < 0.05)$ using the Shapiro-Wilk Test. They also show a heavy distribution of scores towards the higher end, with Prototype, Alpha, Beta, and Gold Phases showing skews of -1.6, -0.69, -0.23, and -1.4, respectively. These indicate that the overall quality of student reflection scores has improved. The improvement in the quality of student reflections is attributed to the increased effort by students due to the graded nature of reflections and the availability of grading rubrics.

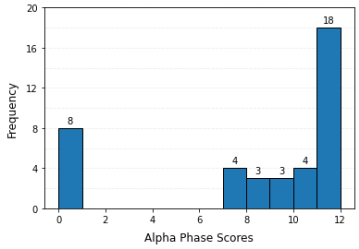
Interestingly, inter-rater reliability has also improved significantly from a poor score of 0.15 to a

Table 9: Fleiss' Kappa for milestone assessments.

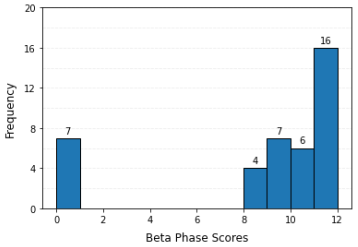
Reflective Score Fleiss' Kappa for CS4350				
Milesstone Assessment	Prototype Phase	Alpha Phase	Beta Phase	Gold Phase
Individual Assessment Kappa	0.49 ($p < 0.05$)	0.61 ($p < 0.05$)	0.55 ($p < 0.05$)	0.56 ($p < 0.05$)
Overall Kappa	0.55 ($p < 0.05$)			



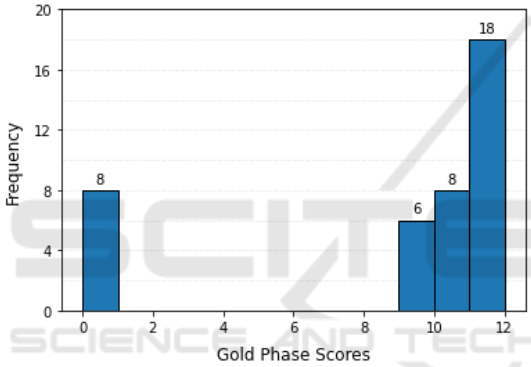
(a) Histogram for Prototype Phase Reflection Scores



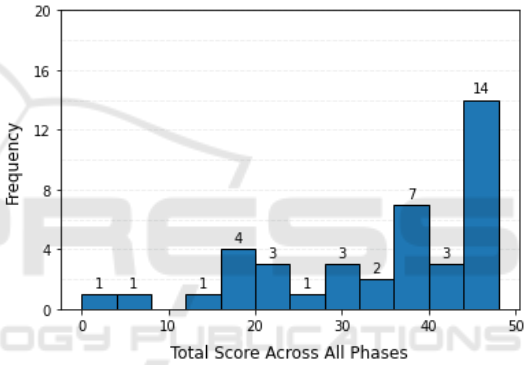
(b) Histogram for Alpha Phase Reflection Scores



(c) Histogram for Beta Phase Reflection Scores



(d) Histogram for Gold Phase Reflection Scores



(e) Histogram for All Phases Reflection Scores

Figure 6: Histogram of CS4350 Phases Reflection scores.

Table 10: Marks achieved by student teams across all phases.

Marks achieved by teams				
Team	**Alpha Phase**	**Beta Phase**	**Gold Phase**	**Final Marks**
1	79.70	88.50	83.80	84.39
2	69.70	95.20	76.65	80.83
3	76.50	86.70	86.65	84.64
4	76.10	78.70	80.30	78.98
5	80.60	98.70	88.05	89.76
6	84.00	96.80	87.05	89.37
7	55.60	32.00	56.30	48.87

moderate score of 0.55. This is likely due to the students tailoring their answers to match what is expected of them as indicated by the rubrics, thus forming more easily identifiable traits for graders to evaluate.

The correlation plot shows significant positive correlations between individual reflection scores and individual marks (specifically Gold Phase individual marks). Mean team reflection scores also generally show a significant positive correlation with marks received by the teams. The exceptions are Beta Phase team marks, which show no significant correlation with any mean reflection scores for any phase, and Team Mean Alpha Phase reflection scores, which show no significant correlation to marks received by teams in any phase. Generally, the correlation between individualized reflection scores and marks is stronger than between aggregated reflection scores and marks. This is likely due to the aggregation process eliminating individual variations that would strengthen the correlation between reflection scores and marks.

In Figure 8, the marks achieved by individuals or teams are plotted against the reflection score, with the line of best fit calculated using ordinary least square regression. The blue region indicates the region with

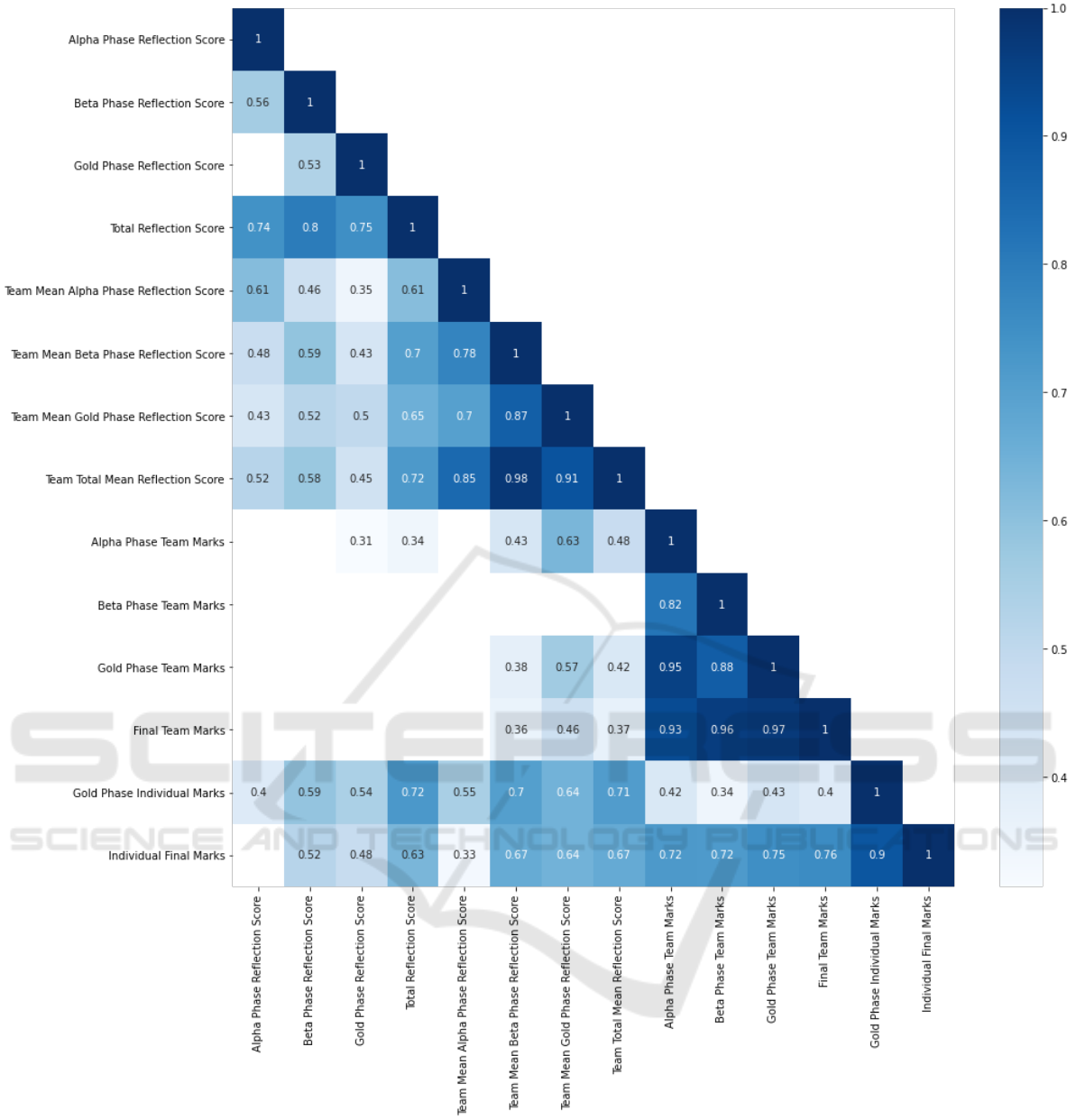
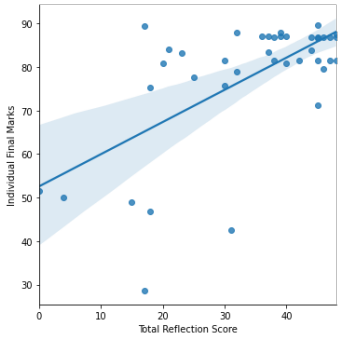


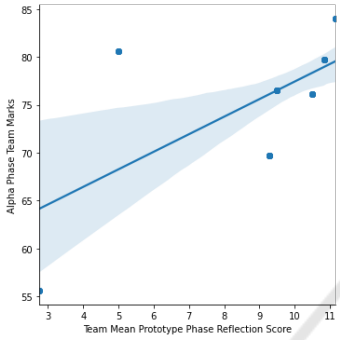
Figure 7: Correlation plot for CS4350 with correlations that are not significant ($p > 0.05$) masked

a confidence interval of 95%. A positive correlation between individual final marks and individual total reflection score is observed in Figure 8a. A positive correlation is also observed for Alpha Phase team marks against the team mean Prototype Phase reflection score in Figure 8b. Gold Phase individual marks against team mean Beta Phase reflection score in Figure 8d also showed a positive correlation. The Beta Phase team marks against the team mean Alpha Phase reflection score in Figure 8c bucks this trend by having a negative correlation. The low number of data points for the aggregated data set could cause this, as any outliers are likely to affect the line of best fit. The single data point with the highest Beta Phase team marks and lowest team mean Alpha Phase reflection score is expected to be the outlier affecting the best-fit line. Overall, it shows that the marks received by students for a phase and their reflection score for the previous phase have a positive correlation.

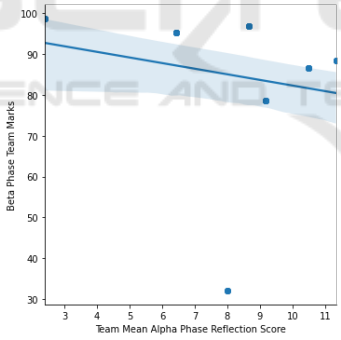
The pre-course survey showed that most students have a generally positive outlook on reflective writing. This might indicate that this batch of students is receptive to reflective writing, contributing to higher-quality reflective submissions. Overall, the percent-



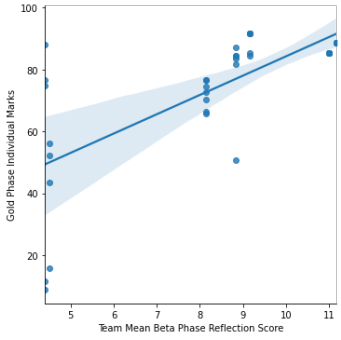
(a) Individual Final marks against Individual Total Reflection Score



(b) Alpha Phase marks against Prototype Phase Reflection Score



(c) Beta Phase marks against Alpha Phase Reflection Score



(d) Gold Phase marks against Beta Phase Reflection Score

Figure 8: CS4350 Marks against Reflection scores.

age of students who found reflective writing useful has decreased slightly from the pre-course survey to the post-course survey. This is attributed to students actually trying out the reflective learning process and finding issues with the current implementation. This can be observed from students' feedback on the reflection writing process. One student suggested a more structured reflection template specific to game development:

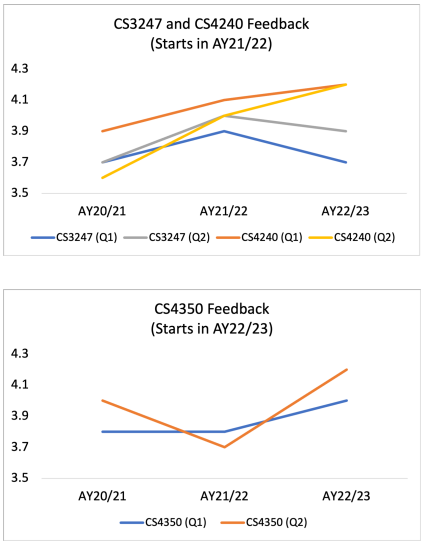
> I really enjoy the reflection writing because it helps me to reorganize my plans and thoughts on further development of my game. However, I think future students could also benefit from a more structured reflection template to review their development process from different perspectives, such as separating design, teamwork, mechanism implementation, and overall level flow designs. It would add more directions for us to learn from.

Another student suggested writing reflections on a group basis instead of individually, "to reflect more on the management of the group and how the management is like at every sprint". Other students suggested providing more guidance on the reflection resources provided and improving the usability of the automated reflection feedback system.

Most students still have a positive impression of reflective writing and find reflections helpful in learning game development. However, the implementation of the reflective learning process can be improved by being more targeted toward collaborative game development, providing more guidance, and increasing the usability of the automated feedback system.

5 ADDITIONAL STUDIES AND DISCUSSIONS

The preliminary study (Section 3) is conducted to understand the difficulties in incorporating reflective learning as formative assessment components in regular computer science and engineering courses. The lessons learned from the preliminary study are used to improve subsequent studies (Section 4). When comparing the overall students' learning experience prior to the introduction of reflective learning, the scores have improved for all three courses, CS3247, CS4240, and CS4350, as shown in Figure 9. However, in the following year (AY22/23), there is a dip in the score for CS3247. This is expected, as we did not use the iReflect tool for CS3247 while keeping the rest of the methods similar to the subsequent study (Section 4). The iReflect tool activities were re-



Q1 - What is your overall opinion of the module?,
Q2 - The effectiveness of teaching method and teaching.

Figure 9: Final Feedback Score.

placed with independent MS-form surveys, and there was no automated feedback for reflection writing. At the same time, CS4240 sores, where we continued to use the iReflect system, have improved further. We conclude that there are challenges in incorporating reflective learning in regular computer science and engineering courses; however, with proper methodologies supported with AI-based tools, it can be successfully implemented to reap the benefits of reflective learning.

5.1 Future Work

Modification of Gibson et al. (2017)'s text analysis pipeline to provide more specific feedback for game development courses will allow students to receive a more tailored response for this course. Based on student feedback, the automated response can improve by providing prompt feedback for game development, such as teamwork, mechanism implementation, and level flow design.

The aggregation of mark data has resulted in a low number of data points, causing outliers to impact correlation and line of best fit significantly. Further studies using unaggregated student data can be done to strengthen the correlation between reflective scores and marks attained.

We are using play-testing and reviewing other teams' work using an adapted version of 100 lenses discussed by Schell (2008) as shown in practice blog (Anand, 2024). Motivated by the benefits of automated feedback for reflection writing in the iReflect system, we are building an AI/ML system that gives automated feedback for play-test reviews for the students to write better peer reviews. This system will be integrated with the iReflect system.

6 CONCLUSION

Based on the preliminary study, no significant relationship existed between reflection scores and students' marks. With the feedback and insights from the preliminary study, a revised methodology was employed in CS4350 to improve the reflection writing process. The change in methodology resulted in an improvement in inter-rater reliability and showed more significant positive correlations between student reflection scores and marks. Despite the additional workload, students also showed a positive attitude toward reflective writing in pre-course and post-course surveys conducted by the research team and the common end-of-the-semester feedback conducted by the University. Given the level of positive responses, we will release our software tool to the community in the near future through software-sharing platforms such as GitHub.

REFERENCES

Anand, B. (2024). Game Analytics with Lenses and Tools. https://www.anuflora.com/game/. [Online; accessed 20-Jan-2024].

Boud, D., Keogh, R., and Walker, D. (1996). Promoting reflection in learning: A model. *Boundaries of adult learning*, 1:32–56.

Bourner, T. (2003). Assessing reflective learning. *Education+ training*.

Cheng, M. W. and Chan, C. K. (2019). An experimental test: Using rubrics for reflective writing to develop reflection. *Studies in Educational Evaluation*, 61:176–182.

Dewey, J. (1933). How we think. a restatement of the relation of reflective thinking to the educative process.

Epp, C. D., Akcayir, G., and Phirangee, K. (2019). Think twice: exploring the effect of reflective practices with peer review on reflective writing and writing quality in computer-science education. *Reflective Practice*, 20(4):533–547.

Gibson, A., Aitken, A., and et al (2017). Reflective writing analytics for actionable feedback. In *Proceedings of the LAK'17*, pages 153–162.

Graesser, A. C., McNamara, D. S., Louwerse, M. M., and Cai, Z. (2004). Coh-metrix: Analysis of text on cohesion and language. *Behavior research methods, instruments, & computers*, 36(2):193–202.

Knight, S., Vijay Mogarkar, R., and et al (2020). Acawriter: A learning analytics tool for formative feedback on academic writing. *Journal of Writing Research*, 12(1):141–186.

Kolb, D. A. (1984). *Experiential learning: Experience as the source of learning and development*. FT press.

Kovanović, V., Joksimović, S., and et al. (2018). Understand students' self-reflections through learning analytics. In *Proceedings of LAK'18*, pages 389–398, USA. ACM.

Mezirow, J. (1991). *Transformative dimensions of adult learning*. ERIC.

Muncy, J. A. (2014). Blogging for reflection: The use of online journals to engage students in reflective learning. *Marketing Education Review*, 24(2):101–114.

Ramsden, P. (2003). *Learning to teach in higher education*. Routledge.

Schell, J. (2008). *The art of game design: a book of lenses*. Morgan Kaufmann Publishers Inc., San Francisco, CA, USA.

Schon (1984). *The Reflective Practitioner: How Professionals Think In Action*. Basic Books.

Schön, D. A. (1987). *Educating the reflective practitioner: Toward a new design for teaching and learning in the professions*. Jossey-Bass.

Sumsion, J. and Fleet, A. (1996). Reflection: can we assess it? should we assess it? *Assessment & Evaluation in Higher Education*, 21(2):121–130.

Tausczik, Y. R. and Pennebaker, J. W. (2010). The psychological meaning of words: Liwc and computerized text analysis methods. *Journal of language and social psychology*, 29(1):24–54.

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(8).

Ullmann, T. D. (2019). Automated analysis of reflection in writing: Validating machine learning approaches. *International Journal of Artificial Intelligence in Education*, 29(2):217–257.

Wong, F., Kember, D., Chung, L., and Yan, L. (1995). Assessing the level of student reflection from reflective journals. *Journal of advanced nursing*, 22(1):48—57.

Yilmaz, F. G. K. and Keser, H. (2016). The impact of reflective thinking activities in e-learning: A critical review of the empirical research. *Computers & Education*, 95:163–173.