Effectiveness of Reflection on Programming Problem Solving Self-Assessments

Mohammed Alzaid, I-Han Hsiao
School of computing, Informatics & Decision System Engineering
Arizona State University
Tempe, AZ, USA
Malzaid1, Sharon.Hsiao@asu.edu

***Abstract* — This Research Work in Progress Paper presents our ongoing work to encourage student to reflect while participating in a distributed self-assessment process. We shed light on the case of low reflect participation and the change in process we made that led to a significant increase in both the quantity and quality of students reflects. To evaluate the effect of the intervention, we conducted two classroom study for an introductory Java course with over 500 students. From the latest enhancements in the current study, we observed that the overall reflecting engagement from the students has significantly increased compared to our initial work. To evaluate the reflects quality, we deployed an automated measure that classifies the constructiveness of the comment based on its relevant and subjectivity to the domain. This work describes the details of the interventions along with the resulted observations. The main objective was to encourage learners to provide thoughtful reflects which we were able to increase along with the positive effect on their performance.**

Keywords—Reflection, Programming learning, self-assessment.

I. INTRODUCTION

The concept of reflection has been agreed upon as a valuable method in the educational settings. It enables the students to make best of their learning process [1][2]. It also provides the ability to the learners to analyze their thoughts, self-motivate, increase their metacognition and strengthen their content knowledge [3]. Reflection may as well enhance the evaluation outcome for the learners when combined with self-assessment [4][5]. While the impact of reflection has been addressed in different domains, it requires more attention from the engineering education research community [6]. For that initiative, we integrated reflection procedure in a home-grown adaptive educational system that provides distributed practices designed for programming novices [7]. The system facilitate distributed reflects as comments provided by the learners during an answer attempt on self-assessment questions. Our objective is to measure the impact of using recommended questions in a personalized programming self-assessment environment as an incentive to leverages students' reflection. Initially we aimed to find the answer for two questions, what triggers the learner engagement in providing reflects during self-assessments? And what effect does distributed reflection has on the overall learner performance in programming courses? Our early work indicated that, reflects during distributed self-assessments had a positive correlation with course performance [7]. This triggered our interest to further explore the impact of reflect quality on the learner's performance as well. However, we were faced with the challenge of low participation from the learners in the reflection procedure along with majorly low-quality comments. We worked on enhancements to find a solution that ultimately would trigger more guided participation from the students. In this paper, we discuss some of the published work related to this research objective, describe the system and the interventions we applied, and present the studies and their preliminary results.

II. RELATED WORK

Because of its known impact on learning process, many researchers attempted to encourage the students to reflect while learning programming courses. Such attempts took different forms and settings, from group activities to self-assessments. In [8], they encouraged reflecting by providing support to the students to submit a weekly plan. The plan shall include their thoughts and evaluation of the knowledge they gained so far. Those efforts also includes encouraging first year computer science students to adopt a better learning strategy by provide their reflection on personal blogs [9][10]. Moreover, the benefits of the reflects can extend the impact on the learner and provide a valuable insight to the instructor. These textual contents can be utilized to become an early indicator that identifies at risk students as well as a predictor of the students' performance. In [11], they performed sequence analysis to predict the course outcome using the reflects from self-assessments and group activities. They reported that their prediction model can achieve 95% accuracy to predict student success in the programming course using the reflects' content. They also observed that, the more the student reflect, the better performance they will get. It's not surprising though, if students put the effort to keep on reflecting for a month, they probably more hard-working students who will end up passing the course. Unlike their work, we are trying to motivate the student to reflect and evaluate the reflects based on its content.

978-1-5386-1174-6/18/$31.00 ©2018 IEEE

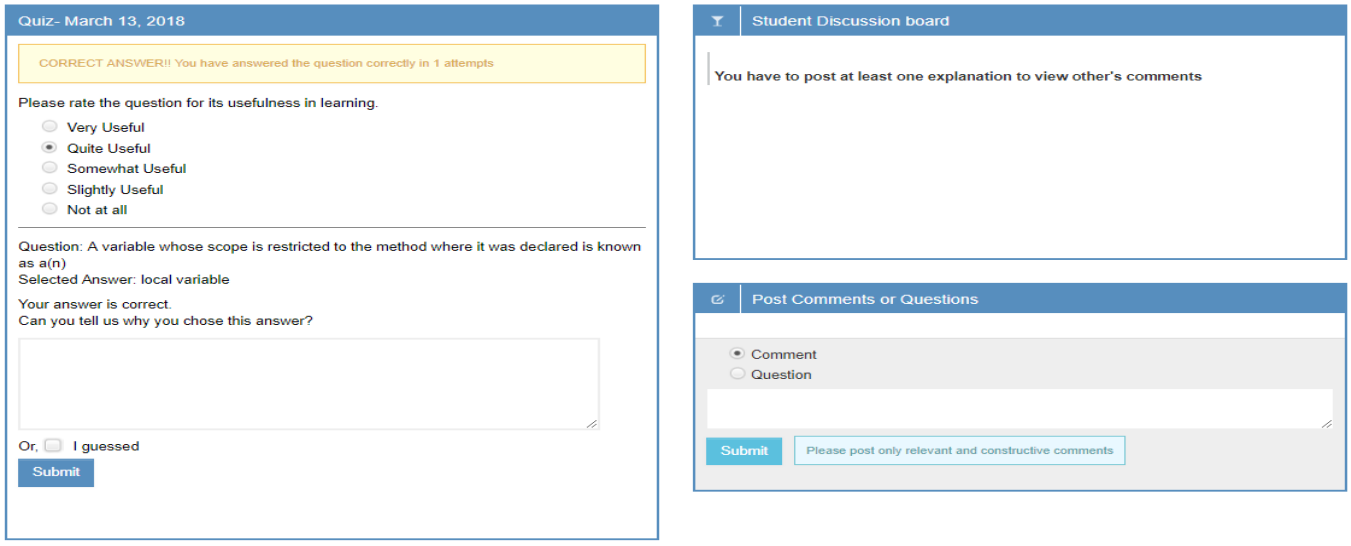

Fig. 1. Answer attempt showing the reflect procedure. On the left, the quiz was successfully answered by the student on the first attempt. The right-side show students' discussion board with no participation from the student.

III. METHODOLOGY

QuizIT is an educational system in an ongoing research project which we continually developing. The system was designed for programming learners to self-assess and monitor their learning progress through daily quizzes. It solicits students' self-explanation as learning reflects while answering a question. Additionally, it supports peer learning opportunities, including reviewing peers' comments, evaluation and annotations on the quizzes. Prior to this work, the system enabled the learners to freely reflect throughout the answer attempts using the designated comment space. When the students decide not to participate in the reflect procedure during the attempt, a notification message will appear stating that "you have not posted any explanation for this question". Viewing other peer's comments in the question discussion board was used as an incentive to trigger more reflects. In the discussion board section on the right in Fig. 1, the student is notified that they can only view this section if they provided at least one explanation or comment.

That feature alone was not enough of a trigger to the students to reflect. Therefore, we considered formalizing the reflect statement [12], adding recommended questions as an incentive [14] and change the flow of the answer attempt to help increase the student engagement in the reflect procedure. We also provided a filtering approach to the discussion board content to only include constructive comments. Next, we describe each of these interventions.

A. Justification prompt

Inspired by the early work in [12], we adopted a similar scaffold approach using justification question to promote reflections. Our assumption was, instead of leaving the student to reflect aimlessly, we would guide them by using a scaffold question. This would help them explain their choices, share their experience which may lead to enhancing the gain from the reflect procedure. The reflect question now will state: "can you tell us why you chose this answer". This question plays the role of directing the reflection towards self-explanation which could improves the student understanding of the question as well [13]. Giving the nature of the multiple-choice questions, the students may prefer to guess the answer when they are uncertain of the correct option. Therefore, we provided the learners with the ability to identify if they had no justification and end up guessing the answer as well. The left side in Fig. 1, shows the reflect procedure after an answer attempt. The student receives the feedback, asked to rate the usefulness of the question and provided the opportunity to reflect using the justification question.

B. Recommended questions

Another incentive we considered to encourage students to reflect was the recommended question feature. The introduction of the recommender was motivated by the students' behavior we saw in a prior study [7]. We observed many learners' decision not to follow up with the daily learning opportunities, with a high turnout occurs around the exam time. To enable the learner, make best of the system, the recommender was introduced to provide them with a personalized experience based on their performance and preferences [14]. The design is based on an open learner model, which can be managed and modified by the learner throughout the semester. The recommender takes into count the student's performance, attempt outcome, question concept, preferences and the peers rating of the questions. The details of the design of the recommender are explained further in [14]. We provided access to the recommended question immediately after the answer attempt once the student submits their comment as an incentive to reflect. If the students skip the reflect procedure, they consequently won't have access to the recommended question on this answer attempt.

Fig. 2. The answer attempt flow of events, the sequence shows how the answer attempt progress from the initial attempt to reattempts.

C. *Change of flow*

The initial design considered enabling consecutive answer attempts, until the correct answer is chosen by the student. This design negatively impacted the reflect procedure and decreased the visibility of the of peers' comments, as students tend to focus more on finding the correct answer before moving to another question. To increase the visibility of the reflect procedure, we modified the answer attempt flow as shown in Fig. 2. Now, the students begin the process with the initial attempt, once they submit their answers, an immediate feedback on the attempt will be provided with the ability to reflect and evaluate the usefulness of the question. Here the students can take a moment to provide their comments and review other peers' comments as well, as seen in Fig. 1. Once the student submits a comment, it will be followed by a personalized recommended question, where only one attempt is allowed. From there, the student may go back to review the question again and decide if they prefer to take another attempt in case the first attempt was incorrect. It's important to mention that, the student may choose not to follow this designated path and move to the dashboard or return immediately to the question list which will prevent them from gaining access to the recommended question feature.

D. *Constructive reflects*

We aim to provide the students with a constructive discussion board content. In earlier versions of QuizIT, this was done by filtering out short reflects based on the number of words in the comment. In deploying a better filtering approach, that identifies the useful comments, we achieve two objectives, maintain the valuable content in the discussion board to make reviewing peer comments more productive. And be able to evaluate the short constructive and distributed reflects effects in programming learning settings. Previously, we attempted to analyze the user generated comments from the answer attempts in QuizIT system. However, we found that the reflects provided in the system did not reach a level of maturity. Since questions tends to be short and precise, the reflects tend to share that feature as well. Reflects are descriptive when they provide evidence of gaining knowledge, by making sense of new experiences and connecting to knowledge source [4]. However, there were no uniform evaluation procedure to assess the quality of the for students' reflection for such short format [15]. Unlike evaluating the comments sentiment in the social media platforms [17], identifying the traces of domain knowledge here makes it more challenging. Therefore, we chose to proceed with the attempt to identify reflections that shows descriptive content as the lowest level of reflection [4]. From analyzing the user generated reflects in QuizIT, we identified the set of comments as descriptive to include subjective and relevant content to the questions. Therefore, we established the rule for the reflect to be descriptive when it shows opinion and present relevant content to the domain of the question. To automate the process of identifying the constructive comments, we began to construct a textual analysis procedure. This step was motivated by the work in [16]. We built the relevant lexicon library based on the subjectivity and relevance to the topics in the questions dataset [17] [18]. As we mentioned earlier, our assumption is that constructive reflects shall provide both opinion and relevant content. We then modified the lexicon to include words from Oracle's Java glossary as per the domain. The comment is labeled relevant, constructive or non-constructive, based on the occurrence of the reflects' words in the lexicon. with a single match, we label the comment as relevant, non-constructive with no matches. When the comment has more than one match with the lexicon we consider it to be a constructive reflect.

IV. EXPERIMENTS

To evaluate the reflect procedure and the following interventions, we performed two classroom studies. The first study was the initial evaluation for the system [7], while the second study evaluated the system enhancements mentioned earlier. To measure the effect of our intervention, we used the same setup in both studies. We introduced the system to an introduction to Java programming course at the beginning of the semesters. Students were encouraged to use the system as non-mandatory tool. In the two-studies the same set of quizzes were scheduled throughout the semester. These questions covered 18 topics (Java, Primitive Data Type, Method, Datatype, Expression, Variables, Strings, Arithmetic's, Operator, Objects, Control, Decisions, Loops, Classes, Constructor, Arrays, 2D Arrays, InputOutput). Three level of question complexity varying from easy, moderate to difficult were provided, and covers conceptual knowledge and programming proficiency by including code and non-code questions.

TABLE 1. DETAILS OF USER ENGAGEMENT FROM THE TWO STUDIES.

Attribute	1st study	2nd study
Registered students	375	152
Number of attempts	11484	5164
Number of comment	247	**988**
Constructive reflects	62	**373**
Constructive users	26	**59**

TABLE 2. SAMPLE OF EVALUATION RESULTS FOR THE CONSTRUCTIVENESS OF USER REFLECTS

Student comments	Constructiveness
I chose "public static void main(**String**[] args)" because the other answers would cause errors.	Constructive
I chose it because of the **semicolon** at the end.	Relevant
I chose this answer because I thought it was correct - it has all the right parts but now I realize the syntax is wrong. There is no space between **main** (**String** [] args)	Constructive
I chose this because it is what we use on all of our codes.	Non-constructive
I do not remember what the exact definition for a **main method** for a **Java** Program is. I have not had my first lab yet so I have had very little practice in Java. I thought it would need the **semicolon** but I think that might be wrong.	Constructive
I forgot there was no **semicolon** at the end, it is suppose to be an open squiggly **bracket**	Constructive

Students were informed by the objective of the study at the beginning of the semester and were given the option to stop at any time.

V. RESULTS

From examining the preliminary data, we observed higher engagement ratio in both the number of attempts and the reflect procedure in the second study, compared to the first study, even though the number of students were much less than the first study participants. In table 1, we can see that after the initial filtering of the reflects, not only the quantity of the reflects has increased but also the quality. In the second study, there were 373 constructive comments out of 988 generated by 59 students, compared to 62 constructive comments generated by 26 students from the first study.

We further analyzed the effect of providing constructive reflects on the student performance. Here, we only consider the first attempt result in study two as the measure of performance. We found that, student who provided constructive reflects performed in their self-assessments, on average (73.1%) success rate, slightly higher than the remaining students who never provided constructive comments (69.2%). Based on the numbers from table 1, we can say that our interventions have made a positive impact and caused more effort from the students. We then examined the impact of the recommender feature; the recommended questions were accessed 73% of the time immediately after the student reflect. This is an indication of the successfulness of the usage of the feature as a reflect trigger. Because without commenting, the students would not be able to immediately access the recommended question. Another observation was the rise in the quality of the reflects. This increase can be linked to the justification question intervention. We noticed that it has a clear impact on the student while formalizing the reflects as can be seen in tables 2. Here, all the reflects are addressing the question as they try to justify their answers. To get a sense of the results of the constructiveness classification approach, table 2 shows the student comments and their labeling results for the following multiple-choice question:

Question: The main method for a Java program is defined by

a) public void main( )

b) public static void main( );

c) private static void main(String[ ] args);

d) public static void main(String[ ] args)

We were also curious to find what is the most discussed concepts in the dataset. We performed textual analysis on the comments and built a visual representation of the result, which we plan to use as another trigger for reflects in the upcoming version of the system.

TABLE 3. THE WORD COUNT FROM THE AGGREGATED STUDENTS REFLECTS.

Word	Count	Word	Count	Word	Count
Class	140	Answer	72	Learned	71
Java	43	Knew	37	Remember	37
Int	36	Lecture	35	Thought	33

In table 3, we can see the top 9 words from the filtered comments. The word that was mentioned the most was Class. This word can refer to the java class or the actual class the students are attending. After examining the data, we found that it mostly referred to the class they are attending, which does align with the observation of student's tendency to reference the source of knowledge after the quiz attempt as can be seen in the table.

VI. CONCLUSION & DISCUSSION

In this work, we presented our effort to encourage the students to reflect during self-assessments. Our approach can be summarized in scaffolding the reflects and providing an incentive to the learners to reflect. To evaluate the effectiveness of the interventions, we conducted classroom studies and presented the preliminary results. We found that after the system enhancements, the usage of the system increased and resulted in higher quality reflects from the students. We also observed that, the performance of the students who provided constructive comments were slightly higher than those who were not able to provide constructive reflects. What we learned from this work was, we can get the students to increase their engagement in self-assessment using simple measures. In the future work, we plan to extend the analysis of the reflects on the student performance and provide other interventions to assist the learners to reflect. We argue that by showing the aggregated comments word cloud to the learners after the answer attempt, it will increase the visibility of the comments and triggers the interest to reflect from the students.

REFERENCES

[1] Dewey, John. How we think. Courier Corporation, 1997.

[2] Bannert, Maria, and Christoph Mengelkamp. "Assessment of metacognitive skills by means of instruction to think aloud and reflect when prompted. Does the verbalisation method affect learning?." Metacognition and Learning 3.1 (2008): 39-58.

[3] Hacker, Douglas J., John Dunlosky, and Arthur C. Graesser, eds. Metacognition in educational theory and practice. Routledge, 1998.

[4] Leijen, Äli, et al. "How to determine the quality of students' reflections?." Studies in Higher Education 37.2 (2012): 203-217.

[5] Thorpe, Mary. "Encouraging students to reflect as part of the assignment process: Student responses and tutor feedback." Active learning in higher education 1.1 (2000): 79-92.

[6] J. Turns, B. Sattler, K. Yasuhara, J. Borgford-Parnell, and C.J. Atman. 2014. Integrating Reflection into Engineering Education.. In Proceedings of the ASEE Annual Conference and Exposition. ACM.

[7] M. Alzaid, D. Trivedi and I. H. Hsiao, "The effects of bite-size distributed practices for programming novices," 2017 IEEE Frontiers in Education Conference (FIE), Indianapolis, IN, 2017, pp. 1-9.

[8] Fekete, Alan, et al. "Supporting reflection in introductory computer science." ACM SIGCSE Bulletin. Vol. 32. No. 1. ACM, 2000.

[9] McDermott, Roger, Garry Brindley, and Gordon Eccleston. "Developing tools to encourage reflection in first year students blogs." Proceedings of the fifteenth annual conference on Innovation and technology in computer science education. ACM, 2010.

[10] Stone, Jeffrey A. "Using reflective blogs for pedagogical feedback in CS1." Proceedings of the 43rd ACM technical symposium on Computer Science Education. ACM, 2012.

[11] Dorodchi, Mohsen, et al. "Reflections are Good!: Analysis of Combination of Grades and Students' Reflections using Learning Analytics." Proceedings of the 49th ACM Technical Symposium on Computer Science Education. ACM, 2018.

[12] Lin, Xiaodong, and James D. Lehman. "Supporting learning of variable control in a computer-based biology environment: Effects of prompting college students to reflect on their own thinking." Journal of research in science teaching 36.7 (1999): 837-858.

[13] Michelene T.H. Chi, Nicholas De Leeuw, Mei-Hung Chiu, and Christian Lavancher. 1994. Eliciting Self-Explanations Improves Understanding. Cognitive Science 18, 3 (1994), pp. 439–477.

[14] M. Alzaid, and I. H. Hsiao, "Personalized Self-Assessing Quizzes in Programming Courses" Proceedings of Workshop on Educational Data Mining in Computer Science Education (CSEDM) in conjunction with The 11th International Conference on Educational Data Mining, 2018.

[15] Dyment, Janet E., and Timothy S. O'Connell. "Assessing the quality of reflection in student journals: A review of the research." Teaching in Higher Education 16.1 (2011): 81-97.

[16] Hsiao, I-Han, and Fatima Naveed. "Identifying learning-inductive content in programming discussion forums." Frontiers in Education Conference (FIE), 2015 IEEE. IEEE, 2015.

[17] Hu, Minqing, and Bing Liu. "Mining and summarizing customer reviews." Proceedings of the tenth ACM SIGKDD international conference on Knowledge discovery and data mining. ACM, 2004.

[18] Wilson, Theresa, Janyce Wiebe, and Paul Hoffmann. "Recognizing contextual polarity in phrase-level sentiment analysis." Proceedings of the conference on human language technology and empirical methods in natural language processing. Association for Computational Linguistics, 2005.