Investigating the Link between Students' Written and Survey-based Reflections in an Engineering Class

Ahmed Ashraf Butt
School of Engineering Education
Purdue University
West Lafayette, USA
butt5@purdue.edu
https://orcid.org/0000-0003-2047-8493

Filiz Demirci
School of Engineering Education
Purdue University
West Lafayette, USA
fdemirci@purdue.edu

Muhsin Menekse
School of Engineering Education and Department of Curriculum and Instruction
Purdue University
West Lafayette, USA
menekse@purdue.edu
https://orcid.org/0000-0002-5547-5455

Abstract— **This study explores the relationship between students' written and survey-based reflections in a first-year engineering class. We collected student reflections using the CourseMIRROR application from 395 students in an engineering class at a midwestern university. After each class during a semester, students were asked to generate a written reflection (in an open-ended format) and their perceived rating (in a Likert-style format) on the lecture's confusing or interesting aspects. We used Spearman correlation statistics to evaluate the relationship between the students' written reflection meta-data (i.e., specificity score and text length) and their perceived lecture rating as confusing or interesting. The results showed that the students tended to rate a lecture as very confusing when they wrote reflections highly relevant to prompts and lecture contents (i.e., reflection quality). Also, we found that the students rating a lecture as very confusing often write a relatively short reflection on the confusing question.**

Keywords— Reflection summary, NLP algorithms, scaffolding, mobile application, learning environment, and reflection quality

I. Introduction

Reflection is considered an effective instructional strategy to foster students' deep learning and enhance their metacognitive skills. It enables students to self-regulate learning by connecting previous knowledge with current learning experiences. It also effectively informs the instructional pedagogy by providing direct feedback on students' learning experiences in the class [1]. Prior studies have shown that reflection writing helps students better understand their learning process [2]. However, integrating reflection activity in large classrooms often proves challenging as it involves manual collection and analysis of students' reflection responses by the instructional team. Therefore, it becomes critical to implement reflection activities that don't overwhelm the students and instructors.

CourseMIRROR is a learning system that facilitates integrating reflection activities in the classroom [3]. This system has two components: 1) website for the instructor and 2) mobile application for the students. The mobile application prompts students to reflect on the lecture's interesting and confusing aspects after the end of each lecture throughout the semester. For each learning aspect, the students were asked to submit a written reflection and rate their interesting or confusing perceptions about the lecture. Students are also provided real-time scaffolding [4] by assessing the reflection specificity during the reflection writing process and guiding users towards writing specific rather than generic reflections.

The reflections are then summarized using Natural Language Processing (NLP) algorithms to provide feedback to the instructors. These reflection summaries are made available to the app and website for the respective user. The system helps to reduce the workload of both instructors and students. Students reflect on their learning from their smart devices in a portable way and get real-time feedback during their reflection writing. For the instructor, the system provides a summary of the students' submitted reflections on the site. This way, instructors can gain insights and feedback from reflections to inform their pedagogy in the practical limitations of time and resources.

This work-in-progress study explores if the reflection meta-data (e.g., reflection specificity score) can be used to improve the summarization part of the learning system. Currently, the summarization algorithm is based on BERTSummExt [5] model, which extends The BERTSumm encoder [5] by adding a sentence classification layer on the top. This encoder modifies the pre-trained BERT model [6] by adding a transformer layer that encodes the complete sentences and their positions. To further enhance the summarization algorithm, the study aims to investigate the relationship between students' written and survey-based reflection in a first-year engineering class. More specifically, this study answers two research questions: 1) What is the relationship between students' reflection quality scores and survey responses? and 2) What is the relationship between students' reflection text lengths and survey responses? The findings of this study will inform our decision about using the students' perceived confusing or interesting aspects as weightage to their written reflection while creating reflection summaries.

II. Related Work

Manual assessment and feedback on reflective writing are key challenges to effectively integrating reflection activities in the classroom [7]–[9]. Traditionally, students' reflection writing was qualitatively analyzed and categorized into common themes to understand the students' difficulties or misunderstandings in the classroom. Also, the majority of the previous work is focused on understanding the depth of students' reflections by analyzing their journals and essays [7]. Furthermore, this traditional approach to analyze the reflection data qualitatively requires a lot of resources in terms of time and human effort, often not available in the classroom. Hence, researchers are exploring ways to generate reflection summaries using computational advancement to facilitate reflection adoption in the classrooms [10].

2023 IEEE Frontiers in Education Conference (FIE) | 979-8-3503-3642-9/23/$31.00 ©2023 IEEE | DOI: 10.1109/FIE58773.2023.10343039

In this regard, researchers are working on automating the process of generating reflection summaries using different Natural Language Processing (NLP) approaches. These approaches include the phrase extraction method [11], MEAD [12], abstractive summarization [13], and LexRnk [14]. Out of these NLP approaches, the abstractive summarization in multitask learning frameworks has produced significant results in producing better reflection summaries using the small training corpus [13]. Different attempts are being made to enhance the abstractive summarization method by integrating it with extractive summarization [15], sentiment classification [16], and text entailment generation [17]. Along the same line, this study is trying to see if the students' perception of a lecture being confusing or interesting can be used as weightage to inform the abstractive summarization while creating reflection summaries. For instance, if the students perceived difficulty rate is aligned with the specificity score of the reflection describing the confusing aspect of the lecture. This information can be used to refine the summary process to create a reflection summary. This involves improving the weightage of reflection that has a strong agreement between the reflection specificity score and their respective rating (either confusing or interesting) of the lecture.

III. Research Design

This study employed a correlational research design to investigate the relationship of the students' reflections meta-data with the students' perceived rating using a CourseMIRROR mobile application in large STEM classrooms.

A. Mobile Educational application

CourseMIRROR (i.e., mobile in-situ reflections and review with optimized rubrics) is a mobile educational application(app) designed to collect the students' reflection [3]. Students can download the application free from the Appstore or Playstore on their smartphone or tablet. The app prompts students to reflect on the confusing or interesting aspects of their learning experiences for each lecture throughout the semester. Fig. 1. shows the reflection questions asked in the reflection activity:

As shown in fig. 1., students were prompted with four questions in each reflection submission. These questions include two open-ended questions to reflect on the confusing reflections (discussing the confusing aspects of the lecture;

a) b) c) d)

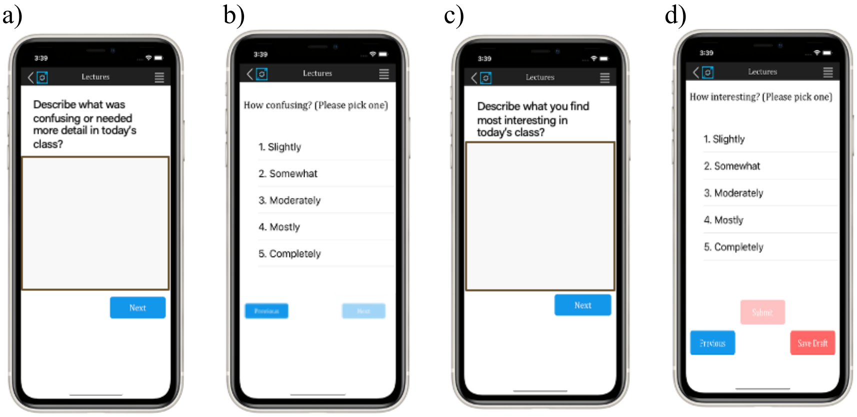

Figure 1. Reflection question in a single reflection submission.

Fig. 1a) and interesting reflections (discussing the interesting aspects of the lecture; Fig. 1c), and two survey-based reflections (their perception of lecture being confusing or interesting; Fig. 1 b & d). Furthermore, the application uses NLP algorithms to provide reflection summaries based on common themes [11]. The summaries are made available on both mobile application and the instructor site. Also, the application scaffold students using the reflection specificity score in their reflective writing process using another set.

B. Site and participants

In this study, we recruited 395 students enrolled in the first-year engineering course at a U.S midwestern university. The topics covered in the course were introductory computer programming concepts, the development of mathematical models, data visualization, and designing solutions for engineering problems. For the study, students voluntarily participated and submitted 9114 reflections in 21 lectures throughout the semester.

C. Reflection specificity model

In this study, we used the NLP algorithm proposed in [18] for evaluating the specificity score of the written reflection. Here, reflection specificity score means the relevancy of the reflection with the reflection prompts and the lecture content. In this approach, the recent SOTA models with a features generation module to evaluate the quality of written reflection text. Furthermore, a classification module was used to assign it a specificity score. Based on the distilled version of the original BERT [19], we used Transformer-based bidirectional deep contextual language models for automatic feature generations. This DistilBERT model reduces the number of parameters to around 60% compared to the original BERT and enables the model to be faster and more suited to the reflection text quality prediction. Furthermore, we used logistic regression classifier to operate on the generated features and produce a quality score of 1, 2, 3, or 4. Furthermore, the logistic regression classification model is trained by keeping the DistilBERT parameters fixed.

D. Data analysis

To inform the study, we used the spearman correlation analysis to investigate the relation among the relationship of the students' reflection specificity score, reflection text length (i.e., number of words) and students perceived rating. Before conducting the analysis, we tested the assumption of the analysis, including the independence of observations and the presence of a monotonic relationship between the variables. We visually inspected to confirm the relationship between variables and confirm the independence of observations by using data collected from different participants.

For analysis, we split the reflection question into two sets: 1) written reflections discussing the interesting aspects and associated students' interesting rating of each lecture, and 2) written reflections discussing the confusing aspects and associated students' confusing rating of each lecture. Our reflection quality model converts all written reflections into an equivalent quality score. For the remaining paper, we will refer to the first and second sets of written reflections text as “Reflections 1.0” and “Reflections 2.0,” respectively.

E. Results

As seen in Table 1, the results indicate that reflection specificity score and text length have a strong positive correlation for both sets of reflection 1.0 (r = 0.735) and reflection 2.0 (r = 0.604). Also, the survey-based reflection has a weak positive relationship (i.e., r = 0.230) with the reflection specificity score of the confusing question and a weak negative correlation with reflection text length.

The result indicated that student who wrote relevant reflections tended to produce lengthier reflection for both question types. Additionally, the students rated lecture highly confusing as they wrote relevant reflection. Overall, this analysis showed that the there is a potential relationship between the students’ perception of a lecture being confusing, and the reflection meta-data.

TABLE 1 SPEARMAN CORRELATION AMONG STUDENTS’ WRITTEN AND SURVEY-BASED REFLECTIONS

	Reflections 1.0		Reflection 2.0	
	1	2	1	2
1. Survey-based reflection	-	-	-	-
2. Reflection specificity score	0.010	-	0.230**	-
3. Reflection text length	- 0.029	0.735**	0.133**	0.604**
**Correlation is significant at the 0.01 level (2-tailed).				

IV. DISCUSSION & CONCLUSION

The current study has explored the relationship between students’ reflection meta-data (i.e., quality score and text length) and their perceived rating of a lecture being confusing or interesting. Our analysis showed that the students writing relevant reflections while explaining their confusion in the lecture often rated it as more confusing. This finding is consistent with the previous literature as it has been suggested that students who engage deeply in the reflection activity become more aware of their learning [2], [20], [21]. Hence, in our case, they were able to rate their perception about the lecture in a better.

Another explanation of the finding could be that the students while reflecting on the confusing aspect of the lecture become more aware of the challenge they faced in the lecture. Hence, they rated the lecture as confusing in a better way. On the other hand, when students reflect on the interesting aspect of the lecture, they didn’t put lot of attention in thinking about the lecture as whole.

The finding of the study contributes to the existing literature trying to automate the analysis of the student’s reflection. Also, this finding will help us to fine-tune our summarization model to create better reflection summaries. As this is work-in-progress study, we will keep exploring the different reflection meta-data alongside the small training corpus to produce better reflection summaries.

These findings should be seen with few limitations in mind. First, we only used the relatively small dataset from single course, thus reducing the generalizability of the finding. Future studies can collect and analyze the students’ reflections from different STEM classes. Second, the study employed correlational design which limits the ability to establish any causal inference. Third, the study used the self-reported measure that is subjected to the participant biases. Hence, future studies can use some objective measure such as physiological measures (e.g., heat rate).

ACKNOWLEDGMENT

The research reported here was supported in part by a grant from the Institute of Education Sciences (R305A180477). The opinions expressed do not represent the views of the U.S. Department of Education.

REFERENCES

[1] I. Boutet, M.-P. Vandette, and S.-C. Valiquette-Tessier, “Evaluating the implementation and effectiveness of reflection writing,” *Can. J. Scholarsh. Teach. Learn.*, vol. 8, no. 1, 2017, doi: https://doi.org/10.5206/cjsotl-rcacea.2017.1.8.

[2] G. van den Boom, F. Paas, J. J. G. van Merriënboer, and T. van Gog, “Reflection prompts and tutor feedback in a web-based learning environment: effects on students’ self-regulated learning competence,” *Comput. Hum. Behav.*, vol. 20, no. 4, pp. 551–567, Jul. 2004, doi: 10.1016/j.chb.2003.10.001.

[3] X. Fan, W. Luo, M. Menekse, D. Litman, and J. Wang, “CourseMIRROR: Enhancing Large Classroom Instructor-Student Interactions via Mobile Interfaces and Natural Language Processing,” in *Proceedings of the 33rd Annual ACM Conference Extended Abstracts on Human Factors in Computing Systems*, in CHI EA ’15. New York, NY, USA: Association for Computing Machinery, 2015, pp. 1473–1478. doi: 10.1145/2702613.2732853.

[4] A. A. Butt, S. Anwar, and M. Menekse, “How do NLP-supported scaffolding techniques support students’ written reflections?,” in *INTED2023 Proceedings*, IATED, 2023, pp. 7450–7450.

[5] Y. Liu and M. Lapata, “Text Summarization with Pretrained Encoders,” in *Proceedings of the 2019 Conference on Empirical Methods in Natural Language Processing and the 9th International Joint Conference on Natural Language Processing (EMNLP-IJCNLP)*, Hong Kong, China: Association for Computational Linguistics, Nov. 2019, pp. 3730–3740. doi: 10.18653/v1/D19-1387.

[6] J. Devlin, M.-W. Chang, K. Lee, and K. Toutanova, “BERT: Pre-training of Deep Bidirectional Transformers for Language Understanding,” in *Proceedings of the 2019 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies, Volume 1 (Long and Short Papers)*, Minneapolis, Minnesota: Association for Computational Linguistics, Jun. 2019, pp. 4171–4186. doi: 10.18653/v1/N19-1423.

[7] V. Kovanović *et al.*, "Understand students' self-reflections through learning analytics," in *Proceedings of the 8th international conference on learning analytics and knowledge*, 2018, pp. 389–398.

[8] M. Liu, S. B. Shum, E. Mantzourani, and C. Lucas, "Evaluating Machine Learning Approaches to Classify Pharmacy Students' Reflective Statements," in *Artificial Intelligence in Education*, S. Isotani, E. Millán, A. Ogan, P. Hastings, B. McLaren, and R. Luckin, Eds., Cham: Springer International Publishing, 2019, pp. 220–230.

[9] A. A. Butt, S. Anwar, A. Magooda, and M. Menekse, "Comparative analysis of the rule-based and machine learning approach for assessing student reflections," in *Proceeding of International Society of the Learning Sciences (ISLS)*, 2022, pp. 1577–1580.

[10] C.-C. Chang, C.-C. Chen, and Y.-H. Chen, "Reflective behaviors under a web-based portfolio assessment environment for high school students in a computer course," *Comput. Educ.*, vol. 58, no. 1, pp. 459–469, Jan. 2012, doi: 10.1016/j.compedu.2011.08.023.

[11] W. Luo, X. Fan, M. Menekse, J. Wang, and D. Litman, *Enhancing Instructor-Student and Student-Student Interactions with Mobile Interfaces and Summarization*. 2015. doi: 10.3115/v1/N15-3004.

[12] D. R. Radev, H. Jing, M. Styś, and D. Tam, "Centroid-based summarization of multiple documents," *Inf. Process. Manag.*, vol. 40, no. 6, pp. 919–938, Nov. 2004, doi: 10.1016/j.ipm.2003.10.006.

[13] A. Magooda, M. Elaraby, and D. Litman, "Exploring Multitask Learning for Low-Resource AbstractiveSummarization," *ArXiv Prepr. ArXiv210908565*, 2021.

[14] G. Erkan and D. R. Radev, "Lexrank: Graph-based lexical centrality as salience in text summarization," *J. Artif. Intell. Res.*, vol. 22, pp. 457–479, 2004.

[15] Y. Chen, Y. Ma, X. Mao, and Q. Li, "Multi-Task Learning for Abstractive and Extractive Summarization," *Data Sci. Eng.*, vol. 4, no. 1, pp. 14–23, Mar. 2019, doi: 10.1007/s41019-019-0087-7.

[16] A. A. Butt, S. Anwar, and M. Menekse, "Work in progress: Uncovering engineering students' sentiments from weekly reflections using natural language processing," in *2023 ASEE Annual Conference & Exposition*, 2023.

[17] R. Pasunuru, H. Guo, and M. Bansal, "Towards Improving Abstractive Summarization via Entailment Generation," in *Proceedings of the Workshop on New Frontiers in Summarization*, Copenhagen, Denmark: Association for Computational Linguistics, Sep. 2017, pp. 27–32. doi: 10.18653/v1/W17-4504.

[18] A. Magooda, D. Litman, A. A. Butt, and M. Menekse, "Improving the quality of students' written reflections using natural language processing: Model design and classroom evaluation," in *International Conference on Artificial Intelligence in Education*, Springer, 2022, pp. 519–525.

[19] J. Devlin, M.-W. Chang, K. Lee, and K. Toutanova, "Bert: Pre-training of deep bidirectional transformers for language understanding," *ArXiv Prepr. ArXiv181004805*, 2018.

[20] M. J. Rodríguez-Triana *et al.*, "Monitoring, awareness and reflection in blended technology enhanced learning: a systematic review," *Int. J. Technol. Enhanc. Learn.*, vol. 9, no. 2–3, pp. 126–150, Jan. 2017, doi: 10.1504/IJTEL.2017.084489.

[21] J. D. Bain, C. Mills, R. Ballantyne, and J. Packer, "Developing Reflection on Practice Through Journal Writing: Impacts of variations in the focus and level of feedback," *Teach. Teach.*, vol. 8, no. 2, pp. 171–196, May 2002, doi: 10.1080/13540600220127368.