Validating the Reflective Writing Framework (RWF) for Assessing Reflective Writing in Computer Science Education through Manual Annotation

Huda Alrashidi[1][0], Mike Joy[1], Thomas Daniel Ullmann[2], Nouf Almujally[1]

[1] University of Warwick, Coventry, UK,[0] h.alrashidi01@gmail.com
[2] The Open University, Milton Keynes, UK

Abstract. The accuracy of a framework for annotating reflective writing can be increased through the evaluation and revision of the annotation scheme to ensure the reliability and validity of the framework. To our knowledge, there is a lack of literature related to the accuracy of any reflective writing framework in Computer Science (CS) education. This paper describes a manual annotation scheme, applied during four pilot studies, to validate the authors' novel Reflective Writing Framework (RWF) for CS education. The results show, through the pilot studies, that the accuracy of Inter-Rater Reliability (IRR) increases from 0.5 to 0.8, which was substantial and close to an almost perfect agreement. This paper contributes to CS education through the reliability and validity of the RWF that can be potentially used for generating an Intelligent Tutoring Systems (ITS) using machine learning algorithms.

Keywords: Reflection, Reflective writing, Computer Science Education, Assessment, Manual Annotation, Framework, Intelligent Tutoring Systems.

1 Introduction

Intelligent Tutoring Systems (ITSs) help to accelerate the student learning process. ITSs are capable of providing timely feedback to students on their writing and so promote students' writing skills. Reflective writing is an essential skill that supports students to improve their knowledge by gaining insight into their personal learning experiences in higher education [1]. Reflective writing has been investigated in Computer Science (CS) education, however, the literature on reflective writing in CS education, that focuses on the investigation of the benefits of reflective writing, is still limited [2, 3].

Few studies have described frameworks which have been adopted or modified from previous studies [3, 4], but these studies did not report the Inter-Rater Reliability (IRR) of those frameworks that have been used to assess students' reflection. In CS education, being able to teach technical skills may be difficult for some as it requires various skills to convey the correct concepts to the students. Reflective writing in the CS education context might be different from other contexts in terms of writing purposes. For example, computer scientists focus on reflection about teamwork, responsibilities for their team role as a coder, communication with complex ideas, keeping skills up-to-date, realising how learning changes over time, and why they may not understanding new

concepts [5]. As a consequence, evaluating such a combination of skills, using manual annotation methods, is challenging [6]. Currently manual assessment through content analysis is used to assess student's reflective writing and is often labour-intensive and time-consuming. For this reason, there is a need for a method that could be used to rapidly assess reflective writing. This study attempts to fill the gap by testing empirically a novel reflective writing framework (RWF) for student writing in CS. The wider goal of this research is to develop an ITS to automate the reflective writing assessment with the provision of timely feedback on students' reflective texts to support the students' learning and to reduce the instructors' efforts on assessment. We focus on the research question: "To what extent can the RWF assess reflective writing in CS"?

2 Methodology

Both qualitative and quantitative methods were used to assess the RWF ability, using an annotated CS dataset, to provide a robust and reliable framework [7], through annotating samples of text that relate specifically to the reflection indicators developed in the RWF. To ensure the RWF validity, face validity was used in this study to ensure the degree to which the indicator categories accurately describe the meaning of characteristics of reflective writing in CS education [8]. To achieve this goal we iteratively reviewed the framework in terms of the reflection levels and indicators. The data used in this work consisted of 60 different reflective writing papers, divided into 760 sentences. The data were gathered from 60 final year CS student projects undertaken during the academic years 2013 through 2016 at the authors' university, and anonymised prior to analysis. The students were asked to write reflectively in terms of contribution, technical achievement, time management, limitation, lessons learned, and future work. These took the students through three levels of questions to lead then to critical reflection: what is the contribution of this project?; why should this project be considered an achievement?; how do you cope when something went wrong?; what are the limitations of this project?; what has been learned?; what would you change if you were to start again? and how can others make use of the work in this project? The IRR, Cohen's k for two raters and Fleiss's k for more than two raters, were used for quality measuring consistency. The raters were recruited based on their experience of assessing formative reflective writing knowledge of reflective writing to produce reliable framework based on the raters' comments and suggestions. The aims were to refine the descriptions of the RWF levels and indicators to ensure the quality of producing annotated reflective writing. These indicators are descriptive, understanding, feelings, reasoning, perspective, new learning, and future action. These relate to the three reflection levels – non-reflective, reflective, and critically reflective – illustrated in Table 1.

Table 1 The RWF with descriptions of reflection levels and indicators (adopted from [9])

Reflection levels	Indicators
Non-Reflective	Descriptive: the writer reports a fact from experience and/or materials Understanding: the writer understands and analyses the experience. Feelings: the writer identifies and analyses their own thoughts and feelings.

Reflective	Reasoning: the writer explains the experience by giving reasons.
	Perspective: the writer shows awareness of alternatives.
Critically-	New learning: the writer integrates and describes new learning
Reflective	Future action: the writer intends to do something and plans for future action.

2.1 Manual Annotation

Once the RWF was developed, a process was undertaken which was aimed at producing a final annotation of the actual dataset created with assistance of the RWF. Four pilot studies were carried out as described in Table 2. Four independent raters were hired to manually annotate 760 sentences in terms of indicators and levels. The raters were required to follow steps: (1) read the sentences; (2) fragment to focus on phrases or words to assess the presence of indicators to determine which level is presented; (3) to rate each sentence according to categorize the sentences that seem to have one or more reflection indicators; and (4) to rate each sentence into one reflective level.

An orientation annotation was conducted to train raters. In this orientation, the raters were introduced to the principle of reflective writing, considering a broad range of possible proposed indicators associated with levels of existing reflection. The focus was to train the raters to be familiar with the RWF. In the first pilot study, 20 sentences were annotated using the RWF by four independent raters. In the second pilot study, three raters were annotated 40 randomly selected sentences. After modifying and retesting the RWF, we further reevaluated and redesigned the RWF guidelines given to the raters. For example, there was a need for such parts to provide sentence examples for each level of reflection and an example of each specific indicator.

Due to the modification of the pilot studies guideline, a consensus was gained regarding three levels and seven indicators in the third and fourth pilot studies. Three independent raters used the RWF and further tested it by using 100 to 400 randomly selected sentences. In conclusion, values of 0.87, 0.78 and 0.80 were achieved for the IRR respectively, which was substantial to an almost perfect agreement [10].

Table 2: The inter-rater reliability computed for each pilot study

Date	#PilotStudy	Sample	#Raters	k
October 2018	1	20	4	0.52
January 2018	2	40	3	0.73
March 2019	3	100	2	0.87
May 2019	4	200	3	0.78
July 2019	4	400	3	0.80

3 Discussion and Conclusions

In this paper, a novel RWF for CS is proposed. This paper answers the research question which focuses on empirically examining the RWF through manual assessment. Our aim was to make the RWF more robust and reliable for calculating the IRR dataset. This emphasizes our wider goal to analyze the quality of reflective writing in

CS education, ultimately to create an ITS to automate reflective writing assessment. According to the results, the quality of the annotated data is assumed to be good since the raters had a similar understanding of the RWF and thus which data should be annotated [11]. This framework can not only help instructors utilize the practice of reflective writing but also provide students with the RWF as a guideline to gain awareness of reflective levels in terms of specific requirements required to reach each level. In future studies, we aim to automate the framework by designing automatic assessment for reflective writing based on rule-based and machine learning algorithms to determine the features of reflective writing samples.

Acknowledgment

The research was partially funded by Kuwait Foundation for the Advancement of Sciences (KFAS) under project code "CB19-68SM-01"

4 References

1. Boud, D., R. Keogh, and D. Walker, *Reflection: Turning experience into learning*. 1985, New York: Nichols Publishing Company.
2. George, S.E., *Learning and the reflective journal in computer science.* Australian Computer Science Communications, 2002. **24**(1): p. 77-86.
3. Chng, S.I., *Incorporating reflection into computing classes: models and challenges.* Reflective Practice, 2018. **19**(3): p. 358-375.
4. Demmans Epp, C., G. Akcayir, and K. Phirangee, *Think twice: exploring the effect of reflective practices with peer review on reflective writing and writing quality in computer-science education.* Reflective Practice, 2019. **20**(4): p. 533-547.
5. Summet, V., *Reflective Writing Through Primary Sources.* The Journal of Computing Sciences in Colleges, 2019: p. 97.
6. Ben-Ari, M., *Constructivism in computer science education.* Journal of Computers in Mathematics and Science Teaching, 2001. **20**(1): p. 45-73.
7. Williams, R.M., et al., *Journal writing to promote reflection by physical therapy students during clinical placements.* Physiotherapy Theory and Practice, 2002. **18**(1): p. 5-15.
8. Krippendorff, K., *Content analysis: An introduction to its methodology*. 2018, CA: Sage publications.
9. Alrashidi, H., M. Joy, and T. Ullmann, *A Reflective Writing Framework for Computing Education*, in *ITiCSE 2019: Proceedings of the 24th Annual ACM Conference on Innovation and Technology in Computer Science Education*. 2019: Aberdeen, UK.
10. Lynn, M.R., *Determination and quantification of content validity.* Nursing research, 1986.
11. Passonneau, R.J. and B. Carpenter, *The benefits of a model of annotation.* Transactions of the Association for Computational Linguistics, 2014. **2**: p. 311-326.