

Self and Socially Shared Regulation of Learning in Data Science Education: A Case Study of “Quantified Self” Project

Jiangxiang Zhang, East China Normal University, zoe_jxiang@163.com
Bian Wu, East China Normal University, bwu@deit.ecnu.edu.cn

Abstract: This study explored the influence of student self-regulation (SRL) ability on socially shared regulation of learning (SSRL) and data literacy presented in a data-driven research project. We adopted a process-oriented method for analyzing video recordings of group conversation in the project meetings. Results showed that the high SRL group were tended to engage more in SSRL and critical components of data literacy than the low SRL group. Implications of the study are also discussed.

Purpose of the study

Our society has become increasingly reliant on data, making it necessary to ensure that all citizens are equipped with the knowledge and skills required to use data effectively (Wolff, Gooch, Montaner, Rashid, & Kortuem, 2016). Regarding the requirement to foster student data literacy development, the emergence of data science as a discipline has only recently been recognized (Kaplan, 2018; Van Der Aalst, 2016). The adoption of data-driven research projects in data science education offers students opportunities to develop skills in , data-oriented communication, data-analytic programming and experiencing uncertainty in data analysis, to name but a few (Kross & Guo, 2019). Productive engagement in this kind of complex problem-solving learning depends on students’ regulation of their individual and shared learning (Järvelä & Hadwin, 2013). However, we are still lack knowledge about the relationship between the regulation of learning and data literacy development in data science education. Therefore, this study aims to explore the influence of student self-regulation ability on socially shared regulation of learning during collaboration on a data-driven research project.

Method

The study was conducted in an undergraduate data science course called “R Programming in Education Application”. To study the impact of SRL on group performance in this data-driven research project, self-regulated learning strategies were measured before group project for assigning students into high or low SRL group by using the motivated strategies for learning questionnaire (MSLQ) (Pintrich, 1991).

The student groups were observed and video-recorded (approximately 270 minutes of recording for each group) as they engaged in collaborative inquiry activities. To investigate the student data literacy from their chat data, we adopted a five-stage framework (Scanlon, Anastopoulou, Kerawalla, & Mulholland, 2011) to develop a data literacy coding scheme, which includes asking questions from data (problem), developing hypotheses and identifying potential sources of data (plan), collecting or acquiring data (data), analyzing and creating explanation from data (analysis), as well as evaluating the validity of explanations based on data and formulating new questions (conclusions). Through reading the transcriptions and viewing the video recordings in iterative cycles, we identified incidents that indicated evidence for cognitive, metacognitive, emotional and motivational regulation processes and then tested and refined the codes into the final coding scheme. Descriptive statistics and Chi Square tests were performed on the coded data, followed by content analysis of conversation excerpts.

Data

The participants of this study were fifty-eight undergraduate students majoring in educational technology at a university in eastern China. The course consisted of six individual learning projects lasting for 90 minutes each and one collaborative project lasting for three weeks with 90 minutes per week. Participants were divided into 10 groups based on their SRL pre-test scores (5 high and 5 low) and worked collaboratively. Each group had of 5-6 members.

Results

To examine how the various SSRL profiles differ in terms of shared cognition, meta-cognition, emotional and motivation, ANOVA was conducted to investigate differences between these ten groups. As the ten groups were divided from a different level of SRL abilities, the results show that students with different SRL levels also have a significant difference in SSRL levels.

We selected two groups with different quality of SRL level (called H-Group and L-Group) to compare their sub-dimensions of SSRL. The results revealed the two groups contrasted in terms of the total number of regulation turns (310 and 272). We compared the two groups' orientation of their cognitive activity, namely, cognitive conflict, information collection and data integration. H-Group spent 12.3% in regulating cognitive activities. On the contrary, the L-Group spent only 3.7% in general. Specifically, H-Group spent the majority time addressing cognitive conflict (7.1 %) compared to only 1.85 % by L-Group. Regarding regulation of meta-cognitive activities, both groups spent around 80% with no significant difference in all three sub-codes, i.e., planning, monitoring and reviewing. As emotional and motivational regulation were concerned, H-Group spent three times more than L-Group for both monitoring (1.9% vs. 0.75%) and control (2.6% vs. 0.75%).

These findings highlight similarities as well as differences in the two groups' SSRL activities with their "Quantified Self" project in their project meeting. Group similarities were found regarding proportion of meta-cognitive activities, which suggests the influence of the task understanding. In contrast, the group differed in terms of cognitive activity. It is noted worthy that H-Group spent more time in setting sub-goals in relation to the task, rather than simply trying to generate their learning objectives. In addition, H-Group expressed more about their emotional and motivational experience, for example encouraging and controlling attention, which contributes to higher efficiency in group collaboration.

Significance of the study

This present study corroborates previous work in suggesting that self-regulation ability is critical predictor of social regulation (Panadero et al., 2015). It also advanced our knowledge in clarifying the relations between regulation learning and data literacy development within a collaboration context of a data-driven research project in data science education.

The theory-driven coding scheme presented in this study was conceptually and methodologically useful to determine how group differences in cognitive activity and metacognitive regulation during project-based collaborative learning contributed to explain differences between the high and low SRL groups' collective understanding. The results found that the high SRL group more easily in developed socially shared regulation of learning compared with the low SRL group. The results also imply that the high SRL group performed better in analyzing and creating explanations from data, as well as evaluating the validity of explanations based on data and formulating new questions, which are two critical components of data literacy.

Some limitations should be taken into accounts when considering our findings. First, despite the advantages of the fine-grained SSRL codes, if we do not have further information regarding how a shared regulatory statement was followed by others to create a joint space of thinking, the analysis is still incomplete. Also, the discourse analysis showed that although it is possible to code several statements of shared regulation within one session, it is also possible not to find an episode of shared regulation as such. This implies that some individual shared statements were ignored or not followed by other group members and may have influence the results. In the future study, we would consider coding self and socially shared regulation from chat data separately and investigating the moderate effect of socially shared regulation between self-regulation ability and data literacy development.

References

- Burke, J., Estrin, D., Hansen, M., Parker, A., Ramanathan, N., Reddy, S., & Srivastava, M. B. (2006). Participatory sensing. Paper presented at the World Sensor Web Workshop, Boulder, CO.
- Cleary, T. J., Callan, G. L., & Zimmerman, B. J. (2012). Assessing Self-Regulation as a Cyclical, Context-Specific Phenomenon: Overview and Analysis of SRL Microanalytic Protocols. *Education Research International*, 2012, 1–19. <https://doi.org/10.1155/2012/428639>
- Gould, R. (2017). Data literacy is statistical literacy. *Statistics Education Research Journal*, 16(1), 22-25.
- Grau, V., & Whitebread, D. (2012). Self and social regulation of learning during collaborative activities in the classroom: The interplay of individual and group cognition. *Learning and Instruction*, 22(6), 401-412.
- Hadwin, A. F., Järvelä, S., & Miller, M. (2011). Self-regulated, co-regulated, and socially shared regulation of learning. In B. J. Zimmerman & D. H. Schunk (Eds.), *Handbook of self-regulation of learning and performance* (pp. 65-84). New York: Routledge.
- Järvelä, S., & Hadwin, A. (2013). New frontiers: Regulating learning in CSCL. *Educational psychologist*, 48(1), 25–39.
- Kaplan, D. (2018). Teaching stats for data science. *The American Statistician*, 72(1), 89-96.
- Kross, S., & Guo, P. J. (2019). Practitioners teaching data science in industry and academia: Expectations, workflows, and challenges. Paper presented at the Proceedings of the 2019 CHI Conference on Human Factors in Computing Systems, Glasgow, UK.