Students' Enactment of Self-Regulated Learning: How does Prior Knowledge Contribute to Learning Traces?

Joni Lämsä, Sanna Järvelä
joni.lamsa@oulu.fi, sanna.jarvela@oulu.fi
Learning and Educational Technology (LET) Research Lab, University of Oulu, Finland

Abstract: The use of digital trace data has enhanced the measurement of self-regulated learning (SRL). However, these learning traces without information about students' internal conditions (e.g., prior domain knowledge) are not sufficient to understand how SRL manifests as a strategic activity. This study investigates how secondary education students' (N = 75) SRL processes differ depending on their prior knowledge on the given source materials while writing from multiple source texts. We used AI to label a set of cognitive and metacognitive processes (e.g., writing, reading, orientation, monitoring) from the trace data and coupled these insights with students' prior domain knowledge test items to label whether a student had high or low prior knowledge of the different source materials. When students engaged with source materials they had high prior knowledge of, they more frequently enacted metacognitive processes together with cognitive ones. We discuss how considering students' internal conditions can enhance the understanding and support of SRL as a strategic activity through human-AI collaboration.

Introduction

Digital trace data has enhanced the measurement of self-regulated learning (SRL) within digital learning environments (Azevedo et al., 2022). For example, Fan et al. (2022) validated the measurement of SRL processes with digital trace data when writing from multiple source texts by combining it with think-aloud data. However, SRL cannot be understood as a strategic activity by merely analyzing the frequencies and durations of patterns derived from trace data (Järvelä & Hadwin, 2024). Instead, self-regulating learners calibrate their SRL processes, for example, based on their prior domain knowledge on the topic they are studying (Winne, 2023). Earlier research has evidenced associations between students' prior domain knowledge and their SRL traces in digital learning environments (Azevedo et al., 2022). When students write from multiple source texts, they likely encounter topics that they have prior domain knowledge of, but also topics where their prior domain knowledge is low (Lämsä et al., 2025). However, only few studies have examined how the enactment of cognitive and metacognitive processes within a single writing task differ depending on whether students are engaging with source material on a topic for which they have high or low prior knowledge. This study aims to address this gap by answering the following research question (RQ): How is students' prior knowledge on the topic of the source materials associated with their enactment of cognitive and metacognitive processes when writing from multiple source texts?

Methodology

Secondary education students (N = 75) learning processes were measured in Finland while they wrote an essay based on multiple source texts in a digital learning environment. The environment included planner, timer, annotation, search, and essay writing tools that made students' cognitive and metacognitive processes traceable during the writing task (Lämsä et al., 2025). Students were given 45 minutes to write a 200-300-word essay on the role of red and white blood cells, using source materials on human biology (Azevedo et al., 2022). Digital trace data included keyboard strokes and navigational activities, such as pages visited or tools accessed. These data were translated into a set of learning actions from which a rule-based AI algorithm (Fan et al., 2022) detected 32 sequences of learning actions that manifested writing, first reading, re-reading, orientation, planning, monitoring, and evaluation processes in the task. Before the task, students filled out a domain knowledge test on the human biology topic (Azevedo et al., 2022). The test included 20 multiple-choice items (α = 0.72), with students scoring an average of 12.5 (SD = 3.1) out of 20 points. The reading materials were divided into 10 sub-sections in the learning environment. The domain knowledge test covered 8 of these sub-sections, with 1 to 4 items per sub-section. We labeled students as having high prior knowledge on the topic of a sub-section if they correctly answered at least 1 out of 1, 2 out of 2, 2 out of 3, or 3 out of 4 items.

To answer the RQ, we applied epistemic network analysis (ENA) using the ENA Web Tool (version 1.7.0) (Marquart et al., 2018; Shaffer et al., 2016). Units of analysis were defined as all lines of data associated with a student engaging with source materials on a topic for which they had high or low prior knowledge (two units per student). The ENA algorithm uses a moving window to construct a network model for each line in the

data, showing how codes (writing, first reading, re-reading, orientation, planning, monitoring, evaluation) in the current line are connected to codes that occur previously. The ENA model was not sensitive to the different window lengths, and we defined it to be 4 lines. Resulting networks were aggregated for each unit of analysis using binary summation, reflecting the presence or absence of code co-occurrence. The model normalized the networks before subjecting them to dimensional reduction, accounting for different numbers of lines across units.

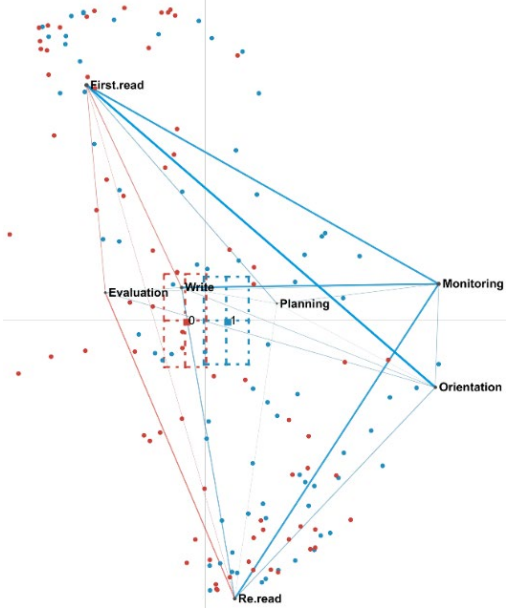
Findings and discussion

The students spent most of their time writing (mean = 59% [SD = 14%]), first-reading (12% [6%]), re-reading (12% [8%]), followed by metacognitive processes of evaluation (7%, [8%]), orientation (7% [4%]), planning (2% [2%]), and monitoring (2% [4%]). These durations did not significantly differ regardless of whether the student engaged with source materials on a topic for which they had high or low prior knowledge. Based on epistemic networks (Figure 1), students enacted more frequently the metacognitive processes of monitoring together with reading and writing, and the metacognitive process of orientation processes together with first-reading when engaging with source materials on a topic for which they had high prior knowledge. When students engaged with source materials on a topic for which they had low prior knowledge, the metacognitive process of evaluation was more frequent. Along the X-axis, showcasing the distinctiveness of the metacognitive processes of monitoring, orientation, and evaluation, the differences when engaging with source materials on a topic for which students had high or low prior knowledge were statistically significant (N=75; t(147.37)= -2.65, p=0.01, Cohen's d=0.43). No differences were found along the Y-axis that was characterized by cognitive processes of reading and writing.

Altogether, our study illustrated how mapping the items from the domain knowledge test with the different source materials enables real-time tracing of whether students are engaging with topics for which they have high or low prior knowledge. Since students' domain knowledge is constantly updating as they engage with the task (Winne & Hadwin, 1998), more fine-grained temporal modeling of students' cognitive and metacognitive processes could be applied in the future. This kind of modeling could be used to advance human-AI collaboration in supporting SRL: For example, generative AI could simplify the source materials on topics where prior knowledge is low (Yan et al., 2024). In this study, students spent relatively the same amount of time engaging with source materials about which they had high and low prior knowledge. Thus, AI could also raise students' metacognitive awareness (Winne & Azevedo, 2022) so that they can be more strategic in allocating their time resources. Future studies should investigate whether students' enactment of cognitive and metacognitive processes while engaging with source materials for which they have high or low prior knowledge improves the predictive power of learning traces in explaining their performance and learning outcomes (Lämsä et al., 2025).

Figure 1

Difference in the Epistemic Networks



The difference in the epistemic networks depending on whether students engaged with source materials on a topic for which they had high prior knowledge (stronger connections in blue) or low prior knowledge (stronger connections in red). There are two dots for each student (N = 75). Dashed lines represent 95% confidence intervals.

References

Azevedo, R., Bouchet, F., Duffy, M., Harley, J., Taub, M., Trevors, G., Cloude, E., Dever, D., Wiedbusch, M., Wortha, F., & Cerezo, R. (2022). Lessons Learned and Future Directions of MetaTutor: Leveraging Multichannel Data to Scaffold Self-Regulated Learning With an Intelligent Tutoring System. *Frontiers in Psychology*, *13*. https://www.frontiersin.org/articles/10.3389/fpsyg.2022.813632

Fan, Y., van der Graaf, J., Lim, L., Raković, M., Singh, S., Kilgour, J., Moore, J., Molenaar, I., Bannert, M., & Gašević, D. (2022). Towards investigating the validity of measurement of self-regulated learning based on trace data. *Metacognition and Learning*, *17*, 949–987. https://doi.org/10.1007/s11409-022-09291-1

Järvelä, S., & Hadwin, A. (2024). Triggers for self-regulated learning: A conceptual framework for advancing multimodal research about SRL. *Learning and Individual Differences*, *115*, 102526. https://doi.org/10.1016/j.lindif.2024.102526

Lämsä, J., de Mooij, S., Aksela, O., Athavale, S., Bistolfi, I., Azevedo, R., Bannert, M., Gasevic, D., Molenaar, I., & Järvelä, S. (2025). Measuring secondary education students' self-regulated learning processes with digital trace data. *Learning and Individual Differences*, *118*, 102625. https://doi.org/10.1016/j.lindif.2024.102625

Marquart, C. L., Hinojosa, C., Swiecki, Z., Eagan, B., & Shaffer, D. W. (2018). *Epistemic network analysis (version 1.7. 0) [Software]* [Computer software]. Available from http://app.epistemicnetwork.org

Shaffer, D. W., Collier, W., & Ruis, A. R. (2016). A Tutorial on Epistemic Network Analysis: Analyzing the Structure of Connections in Cognitive, Social, and Interaction Data. *Journal of Learning Analytics*, *3*(3), 9–45. https://doi.org/10.18608/jla.2016.33.3

Winne, P. H. (2023). Roles for Information in Trace Data Used to Model Self-Regulated Learning. In V. Kovanovic, R. Azevedo, D. C. Gibson, & D. lfenthaler (Eds.), *Unobtrusive Observations of Learning in Digital Environments: Examining Behavior, Cognition, Emotion, Metacognition and Social Processes Using Learning Analytics* (pp. 175–196). Springer International Publishing. https://doi.org/10.1007/978-3-031-30992-2_11

Winne, P. H., & Azevedo, R. (2022). Metacognition and Self-Regulated Learning. In R. K. Sawyer (Ed.), *The Cambridge Handbook of the Learning Sciences* (3rd ed., pp. 93–113). Cambridge University Press. https://doi.org/10.1017/9781108888295.007

Winne, P. H., & Hadwin, A. F. (1998). Studying as self-regulated learning. In D. Hacker J., J. Dunlosky, & A. C. Graesser (Eds.), *Metacognition in educational theory and practice* (pp. 277–304). Lawrence Erlbaum Associates Publishers.

Yan, L., Greiff, S., Teuber, Z., & Gašević, D. (2024). Promises and challenges of generative artificial intelligence for human learning. *Nature Human Behaviour*, *8*(10), 1839–1850. https://doi.org/10.1038/s41562-024-02004-5

Acknowledgments

The work was supported by funding from the Jacobs Foundation (the Center for Learning and Living with AI, CELLA) and Research Council of Finland (grant numbers 361208 and 324381).