Triangulating Multimodal Data to Measure Self-Regulated Learning

Author(s): Lim[1], K.P., van der Graaf[2], J., Fan[3], Y., Engelmann[1], K., Bannert[1], M., Molenaar[2], I., Gasevic[3], D., Moore[3], J.
Technical University of Munich[1]
Radboud University[2]
University of Edinburgh[3]
lyn.lim@tum.de

ABSTRACT: The poster presents the first study of a project which focus on improving measurement and real-time support of self-regulated learning (SRL) using learning analytics. Education is increasingly focused on students' ability to regulate their own learning within technology-enhanced learning environments. Current SRL interventions do not sufficiently adapt to the individual learning process, thus, learning analytics offer an approach to better understand SRL-processes. As current approaches lack validity or require extensive analysis after the learning process, we aim to investigate how to advance support given to students by 1) improving unobtrusive data collection and machine learning techniques to gain better measurement and understanding of SRL-processes and 2) using these new insights to facilitate students' SRL by providing personalized scaffolds. We will reach this goal with a series of exploratory, lab, and field studies. The setup presented here consisted of a learning environment presented on a computer with a screen-based eye-tracker. Other data sources are log files, screen recording, and audio of students' think aloud. The analysis will focus on aligning the different data sources and detecting sequences that are indicative of micro-level SRL processes as a stepping stone for improving real-time scaffolds.

Keywords: self-regulated learning; personalized scaffolds; learning analytics; machine learning; adaptive systems

1 TRACE DATA TO DETECT SELF-REGULATED LEARNING PROCESSES

This project (funded by ORA; BA20144/10-1, NWO 464.18.104, ES/S015701/1) aims to improve measurement of self-regulated learning (SRL) by using multimodal learning analytics. SRL occurs when learners monitor and regulate content they access and operations they apply to operate on content as they pursue goals to augment and edit prior knowledge (Winne, 2019). Previous studies have shown that SRL is related to better learning outcomes and interventions can improve SRL and learning outcomes (e.g. Bannert & Reimann, 2012). However, the need for improved SRL measures has increased to capture processes while they occur (Schunk & Greene, 2018) as there is still no agreement to the appropriate learning actions to measure, diagnose, understand, and support students' SRL (Papamitsiou & Economides, 2014). A solution is to assess SRL at a more fine-grained level by measuring micro-level SRL processes. Unobtrusive measures of SRL can be captured through trace data in digital learning environments. Such traces are less biased than self-reports due to their temporal proximity (Gasevic, Jovanovic, Pardo, & Dawson, 2017), but traces do not reflect SRL processes on their own (Molenaar & Järvelä, 2014). Think aloud data has shown to be more insightful in determining SRL activities and predicting students' learning achievements than self-reports (Bannert, 2007). The integration of trace data with think aloud data provides opportunities to better measure micro-level SRL processes.

2 MEASURING MICRO-LEVEL SRL PROCESSES

This project will investigate and improve log data in two exploratory studies and develop and test personalized scaffolds based on individual learning processes in two laboratory and one subsequent field study. All studies are set in a scenario in which students are tasked to learn about artificial intelligence, differentiation, and scaffolding, and to write an essay, see Fig. 1 (left) for a schematic overview of the exploratory studies. Before and after this task, students' knowledge about the topics is assessed. Preliminary results of the first study show that there is a significant learning gain. How this learning gain is related to micro-level SRL processes will be investigated in the exploratory studies. Three types of data will be gathered: think aloud (audio), log data (mouse and keyboard), and eye-tracking (gaze).

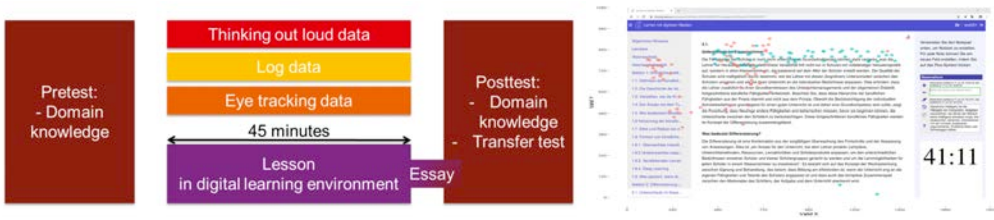

Figure 1: Study design (left) and the digital learning environment with eye fixations (right).

Think aloud can be coded to detect cognitive and metacognitive processes (micro-level SRL processes). But it is unclear if and how this process presents itself in log data or eye-tracking, see Fig. 1 (right). Think aloud data has been shown to be a good indicator of SRL processes and past research has paved a clear path for its analysis. The relation between SRL processes and the other data sources, however, is less evident. To tackle this, think aloud data will be used as an indicator of SRL and other data sources as proxies of SRL processes. In particular, sequential patterns are expected to be indicative of SRL processes. The challenge is to triangulate individual data streams and find proxies of micro-level SLR processes in log data.

2.1 Experimental setup and data analysis approach

The following setup were used: Screen-based eye-trackers (Tobii Pro Spectrum/TX300), webcams with microphones, keyboards, and mice. Data were collected synchronously on a computer using iMotions/Tobii Studio while the learning environment was presented. Multiple data sources were configured and combined into a log file. *Audio* was recorded to measure think aloud data—used for coding SRL processes. Previously developed and validated coding schemes for SRL were used to score the think aloud data (Bannert, 2007). *Log data* (mouse and *keyboard* data*)* indicate how the participant interacted with the learning environment. *Eye tracking data* was sampled at 300 Hz and consisted of fixations, saccades, gaze points, pupil size etc.

The data analysis was conducted in two steps: (a) We developed a trace parser which processed the trace data by first labeling all raw trace data to form an action library. A collection of actions formed meaningful learning patterns (i.e. a pattern library). These patterns formed the overall categories (e.g. Planning) in order to map to the think aloud data. (b) Next steps include the combination of multiple data streams.

2.2 Preliminary results

We segmented and coded think aloud protocols. To demonstrate the alignment of data sources, periods of orientation (metacognition) and reading (cognition) were contrasted, see Table 1. The difference in codes for the think aloud and log files showed that orientation was related to interactions with the menu (i.e. navigation), while reading did not leave any traces in the log data. For eye-tracking, the distance between subsequent fixations was larger for orientation (M = 105.57, SD = 65.84) compared to reading (M = 79.12, SD = 41.16), $t(110)$ = 2.54, p = .013, d = 0.48.

Table 1: Triangulating multimodal data of SRL events to think aloud

Data source	Observations and possible processes from multiple data streams	
Think aloud	Orientation (*Task analysis)*	Reading (*Reading content)*
Log files	Navigation (*Task overview)*	None *(Accessing content)*
Eye tracking	Large distance (*Overview*)	Small distance (*Details*)

3 DISCUSSION AND NEXT STEPS

Using think aloud data to shed light on the measurement of SRL processes with multimodal data facilitates our understanding on how, when, and for whom to provide real-time support during SRL. In the example above, navigation actions and fixation patterns in eye-tracking data appeared to be an indicator for orientation. This shows how learning analytics can be applied in SRL research to better understand the learning process and ultimately, support SRL. The results are a stepping stone that demonstrate how meaning can be uncovered in multimodal data through the use of think aloud data as the ground truth. The project continues by developing an advanced algorithm to analyze learning processes and test its application in authentic learning settings.

REFERENCES

Bannert, M. (2007). *Metakognition beim Lernen mit Hypermedia*. Münster, Germany: Waxmann

Bannert, M., & Reimann, P. (2012). Supporting self-regulated hypermedia learning through prompts. *Instructional Science, 40*, 193–211.

Gasevic, D., Jovanovic, J., Pardo, A., & Dawson, S. (2017). Detecting Learning Strategies with Analytics: Links with Self-reported Measures and Academic Performance. *Journal of Learning Analytics*, *4*(2), 113–128–113–128.

Molenaar, I., & Järvelä, S. (2014). Sequential and temporal characteristics of self and socially regulated learning. *Metacognition and Learning*, *9*(2), 75–85.

Papamitsiou, Z., & Economides, A. A. (2014). Learning analytics and educational data mining in practice: A systematic literature review of empirical evidence. *Journal of Educational Technology & Society, 17*, 49-64.

Schunk D.H., & Greene J.A. (2018). *Handbook of Self-Regulation of Learning and Performance (2nd ed.)*. London, UK: Routledge.

Winne, P.H. (2019). Paradigmatic Dimensions of Instrumentation and Analytic Methods in Research on Self-Regulated Learning. *Computers in Human Behavior, 96*, 285-289.