Building theory-informed learning analytics to understand and intervene in Socially-Shared Regulation of Learning

Cristina Villa-Torrano
GSIC-EMIC Research Group, Universidad de Valladolid, Spain
cristina@gsic.uva.es

ABSTRACT: There is a need to strengthen regulatory processes in collaborative learning. The Socially-Shared Regulation of Learning (SSRL) theory aims at understanding the regulatory processes through which group members negotiate objectives, planning, and strategies for carrying out a collaborative activity. Some studies on this topic have been conducted using students' self-reported or physiological data. However, self-reported data is biased by the students' perception and invasive sensors are costly and cumbersome. Moreover, these studies do not provide actionable information on time. Additionally, the analysis of SSRL becomes even more challenging when not restricted to specific learning environments or learning situations. Therefore, we propose to use Learning Design to guide data collection and inform the learning analytics using trace data from different technological tools. Through this, we expect to build predictive models that provide actionable information on SSRL, with a methodology that is not restricted to a specific learning design.

Keywords: Socially-Shared Regulation of Learning, Collaborative Learning, Learning Analytics, Learning Design.

1 BACKGROUND

Collaboration is one of the 21st Century Skills (Voogt & Roblin, 2010) that is increasingly present in academic and work context (Malmberg et al., 2015). Collaborating with others benefits learning yet comes with some challenges (Kreijns et al., 2003) that students need to overcome with their peers to achieve the shared learning goals (Malmberg et al., 2015). As noted by (Järvelä et al., 2020), success in collaborative learning often occurs when team members systematically activate and maintain their cognition, motivation, and emotions towards the achievement of their shared goals, i.e., socially regulating team efforts. Moreover, many empirical studies show that regulatory processes are critical for the success of collaborative learning (Järvelä et al., 2016).

Socially-Shared Regulation of Learning (SSRL) is a field in the framework of self-regulated learning theories that integrates different types of collective regulatory processes that contribute to shared regulation (Hadwin et al., 2011). Shared regulation processes happen when team members negotiate the perception of tasks, objectives, planning, and strategies. SSRL is theorized to consist of four stages that are interconnected and can be recursive (Malmberg et al., 2015): i) negotiation and construction of the perception of the task, based on internal and external representations; ii) sharing of objectives and generating plans to achieve them; iii) coordination and monitoring of progress; iv) reflection and redesign of objectives, planning or perception of activities. There exists initial evidence that successful groups are those that use multiple regulatory processes; students start using self-regulatory processes, such as task understanding and monitoring, and then perform shared regulation processes, such as jointly making plans for how to approach the task (Malmberg et al., 2015).

SSRL has already been explored from several perspectives. There are works where SSRL is studied using self-reported data about the challenges perceived by the groups and analyzing what SSRL strategies they develop to overcome them (Malmberg et al., 2015). In other works, such as (Malmberg et al., 2017), groups collaboratively carry out an assignment and then have to answer a questionnaire related to shared understanding, challenges, planning, etc. In this case, SSRL is studied through the conversations that students have through an online platform. Recently, SSRL has been researched by analyzing physiological data, observation data (video) and expression recognition (Järvelä et al., 2019). Three main limitations can be identified in these works: i) the validity of their findings is limited to very specific platforms and learning situations, and they do not consider how the pedagogical design and intentions shape the collaborative behavior and relevant regulatory processes; ii) the data was obtained through self-reported instruments or invasive sensors. However, the literature shows that students are biased when asked what regulatory processes they have followed (Saint et al., 2020), while the use of invasive sensors is less likely to be widely accepted; and iii) the focus of these studies has been on understanding regulatory processes post-hoc, but not in supporting these regulatory processes with actionable information for teachers and/or students during the enactment of the learning situation.

Previous work has shown that one way of including contextual and pedagogical information in the analysis is by means of the Learning Design (LD) (Rodríguez-Triana et al., 2015). Over the last two decades, the LD research field has been proposing processes and tools aimed at effectively supporting the complex task of conceptualizing and elaborating activity plans that can be enacted, shared and repurposed (Conole, 2013; Mor & Craft, 2012). Previous works suggest that LD can help in collecting learning data, in making meaning out of it, and in analyzing it (Lockyer et al., 2013). Therefore, and if we do not want to propose ways of analyzing SSRL processes that are restricted to concrete and specific learning situations, we need to propose LA approaches for SSRL that can be applied to learning environments in which different learning designs can be supported and enacted. In such LA approaches to SSRL, LD would be expected to play a significant role. Virtual Learning Environments, Distributed Learning Environments, and even MOOC platforms are examples of such environments that can support different learning situations and that, at the same time, provide trace data about students' behavior. Using this trace data, instead of (more biased) self-reported data and/or (difficult to collect) physiological data, we expect, on the one hand, to be able to understand shared regulation processes and, on the other hand, to be able to detect optimal and sub-optimal patterns of shared regulation during the different phases that are expected to happen, according to the learning design, to be able to make early interventions. For the latter, we expect to build predictive models based on the optimal and sub-optimal processes detected in order to provide actionable indicators.

Some of these features are being pursued by current research in the area of self-regulated learning (Jovanovic et al., 2020; Saint et al., 2020), where the traces of online tools are coded into macro-level constructs (e.g., planning) which comprise micro-level actions (e.g., setting goals or making personal plans) based on the theoretical models of SRL. In these studies, researchers detect predictable patterns that could inform the development of automatic interventions to provide real-time feedback (Saint et al., 2020). In this thesis, we follow a similar approach where the learning designs would help identifying macro-level constructs identified by SSRL theories and the micro-level actions would correspond to the students' actions. We expect to detect optimal and sub-optimal patterns of shared regulation from the traces of online platforms and that these patterns can help to create predictive models with actionable information. To the best of our knowledge, this approach has not been followed in the SSRL literature.

2 RESEARCH QUESTIONS AND GOALS

The underlying research question of this doctoral thesis is: **How can Learning Analytics based on the theory of SSRL help identify and predict patterns of shared regulation that provide actionable information in collaborative learning situations using trace data?** Our approach to answer this question is to automatically extract meaningful features from trace data considering the learning design that defines the collaborative learning situation in which SSRL processes are expected to happen. The general objective (to provide actionable information by detecting patterns of SSRL using trace data) is divided into two particular objectives:

1. **To map event data to SSRL theory constructs.**
   According to (Siadaty et al., 2016), precise conceptual SSRL models need to be defined. Based on a specification of SSRL constructs, we will explore how teachers can be involved to inform learning designs with additional information about where and when regulatory processes are expected to happen. This will help match trace data produced in the enactment of the activities to the appropriate SSRL construct. The mapping of traces to SSRL phases/constructs can help to identify important features to detect shared regulation patterns.

2. **To provide actionable information by building early predictive models of successful collaboration based on SSRL patterns.**
   It should be explored which learning analytics techniques could detect optimal and sub-optimal SSRL patterns through the mapped data. Once the above objective is achieved, it will be possible to identify which features can help to make early predictions. Furthermore, the detection of optimal and sub-optimal SSRL patterns would provide actionable information for early interventions.

3 BRIEF STATE OF THE ART

In recent years, a number of empirical studies have been conducted in the area of SSRL. In particular, a learning environment with regulation tools was used in (Malmberg et al., 2015) to prompt students to recognize challenges that may hinder collaboration and to develop SSRL strategies to overcome them. This study employs students' self-reported answers to the questions asked in the virtual environment, coded by the authors. The result of this research indicates that there is a difference between the regulatory processes followed by high and low performing groups. On the other hand, (Malmberg et al., 2017) focuses on the temporal and sequential order of the different types of regulation (self-regulation, co-regulation and socially shared regulation of learning) in collaborative activities. The data used in the study consists of videos of the working groups during two months in a math didactics course. Finally, in (Järvelä et al., 2019), a preliminary study uses data from different sources to help understand SSRL processes. Specifically, the use of physiological sensors is explored in greater depth, as is also detailed in (Järvelä et al., 2020).

These studies have been carried out with self-reported data or physiological data from students using invasive sensors. However, regulation can also be mapped to dynamic series of events that change over the learning situation (Siadaty et al., 2016) using traces from learning platforms. Furthermore, the studies mentioned on SSRL focus on understanding the processes of shared regulation, but not on making early predictions that allow timely interventions. This approach has recently started to be researched in the area of Self-Regulated Learning (SRL) through process mining (Saint et al., 2020). These works detect predictable patterns that could provide actionable information to trigger feedback in real time. However, to the best of our knowledge, it has not been researched in the area of SSRL. Moreover, these studies do not consider the context and pedagogical intentions behind each activity of a learning situation. As we mentioned before, regulation and social processes change along the learning situation (Malmberg et al., 2015). Therefore, it is important to align the learning design and learning analytics in order to: i) inform about the processes that are expected to occur during the situation (Er et al., 2019); and ii) guide the collection of data and the analysis to be made. Although the connection between learning design and learning analytics is growing significantly in the literature (Lockyer et al., 2013; Rodríguez-Triana et al., 2015), to the best of our knowledge, it is not being considered in the area of SSRL.

4 METHODOLOGY

The proposed methodology to answer the research question is Design Science Research Methodology (DSRM) (Peffers et al., 2007). DSRM aims at the creation and evaluation of artifacts that solve problems, like constructs, models or any designed object that offers a solution to the research problem. This methodology defines a process model involving the following phases: (i) identify a problem and motivate its interest; (ii) define the objectives of a solution; (iii) design and develop an artifact for the solution; (iv) demonstrate how the artifact solves the problem; (v) evaluate it; and (vi) communicate its performance. These phases do not need to happen necessarily sequentially. Indeed, refinements of the proposed solutions are foreseen by iteration through the different activities.

The overarching objectives of this thesis and its iterative nature make DSRM a suitable methodology to frame this thesis work. This PhD thesis aims to design and develop artifacts that provide actionable information by detecting patterns of SSRL using trace data. During the thesis, we need to involve the main stakeholders (teachers, learning/instructional designers, students, …) with several purposes, including: identify and describe learning scenarios that can benefit from SSRL, explore how teachers can be involved to inform learning designs with additional information about where and when regulatory processes are expected to happen, evaluate the degree in which the solutions meet the needs of the participants, etc.

Regarding the number of iterations needed, we foresee three iterations. The first iteration consists of a literature review focusing on theoretical models and the adoption of these models in empirical studies to support collaboration. This literature review is complemented with an exploration of the relevant data sources, machine learning techniques and actionable information to generate in relation to SSRL. Moreover, a first conceptual solution is proposed, and it is evaluated by exploratory studies, that will help in turn to understand better the problem and the goals. During the second iteration, the conceptual and technological solution to solve the detected gaps will be refined and developed. It is expected that, during this iteration, two studies can be conducted to evaluate part of the proposal. Finally, in the third iteration the proposal will be improved with the previous evaluations and the final evaluation of the proposed solution will be carried out.

5 CURRENT PROGRESS

So far, the author has been working on the first iteration of the thesis plan. She has carried out a non-systematic review of the state of the art of SSRL and SRL, focusing on the definition of the theoretical models, the adoption of these models in empirical studies and the types of data collected. In addition, the author, and her colleagues have submitted a paper to an international conference where they work in an exploratory collaborative scenario where they detect shared regulation processes through trace data. The collected data was coded based on the theoretical model and they detect SSRL processes using a process mining technique. The theory-informed LA also helped to interpret the processes of shared regulation and to detect behavior that was not expected during the activity. However, this study was conducted using data coming from an online learning platform designed to support a specific type of collaborative activity and the learning design was very concrete. Furthermore, this exploratory scenario has helped us to identify different aspects of a specific collaborative learning scenario: regulation processes that occur, data that can be collected, interventions that can be made, … As a result, it will facilitate the definition of the scenarios that are relevant for answering our research question.

Since our main objective is to provide actionable information by detecting patterns of SSRL using trace data, the next steps are: i) to identify and describe additional collaborative scenarios that illustrate how teachers can benefit from our approach. ii) to identify which types of data sources can help us detect SSRL patterns, as suggested by the learning design of previously detected scenarios; iii) to identify what actionable information we want to generate through SSRL; iv) to explore the use of machine learning techniques (e.g., process mining) to discover SSRL patterns. Then, we have to put them into practice with accessible datasets. It is expected that we will be able to conduct two studies during the second term of this academic year, where we expect to detect SSRL patterns through both different platforms and learning designs.

ACKNOWLEDGEMENTS

This research is partially funded by the European Regional Development Fund and the National Research Agency of the Spanish Ministry of Science, Innovations and Universities under project grants TIN2017-85179-C3-2-R.

REFERENCES

Conole, G. (2013). Designing for learning in an open world. In *Designing for Learning in an Open World*. https://doi.org/10.1007/978-1-4419-8517-0

Er, E., Gómez-Sánchez, E., Dimitriadis, Y., Bote-Lorenzo, M. L., Asensio-Pérez, J. I., & Álvarez-Álvarez, S. (2019). Aligning learning design and learning analytics through instructor involvement: a

MOOC case study. *Interactive Learning Environments*. https://doi.org/10.1080/10494820.2019.1610455

Hadwin, A. F., Järvelä, S., & Miller, M. (2011). *Handbook of self-regulation of learning and performance* (pp. 65–84). Educational psychology handbook series.

Järvelä, S., Gasevic, D., Seppänen, T., Pechenizkiy, M., & Kirschner, P. A. (2020). Bridging learning sciences, machine learning and affective computing for understanding cognition and affect in collaborative learning. *British Journal of Educational Technology*. https://doi.org/10.1111/bjet.12917

Järvelä, S., Malmberg, J., Haataja, E., Sobocinski, M., & Kirschner, P. (2019). What multimodal data can tell us about the students' regulation of their learning process? *Learning and Instruction*. https://doi.org/10.1016/j.learninstruc.2019.04.004

Järvelä, S., Malmberg, J., & Koivuniemi, M. (2016). Recognizing socially shared regulation by using the temporal sequences of online chat and logs in CSCL. *Learning and Instruction*. https://doi.org/10.1016/j.learninstruc.2015.10.006

Jovanovic, J., Dawson, S., Joksimovic, S., & Siemens, G. (2020). Supporting actionable intelligence: reframing the analysis of observed study strategies. *Companion Proceedings 10th International of Conference on Learning Analytics & Knowledge (LAK 2020)*, 161–170. https://doi.org/10.1145/3375462.3375474

Kreijns, K., Kirschner, P., & Jochems, W. (2003). Identifying the Pitfalls for Social Interaction in Computer-Supported Collaborative Learning Environments: A Review of the Research. *Computers in Human Behavior*, *19*, 335–353. https://doi.org/10.1016/S0747-5632(02)00057-2

Lockyer, L., Heathcote, E., & Dawson, S. (2013). Informing Pedagogical Action: Aligning Learning Analytics With Learning Design. *American Behavioral Scientist*. https://doi.org/10.1177/0002764213479367

Malmberg, J., Järvelä, S., & Järvenoja, H. (2017). Capturing temporal and sequential patterns of self-, co-, and socially shared regulation in the context of collaborative learning. *Contemporary Educational Psychology*, *49*. https://doi.org/10.1016/j.cedpsych.2017.01.009

Malmberg, J., Järvelä, S., Järvenoja, H., & Panadero, E. (2015). Promoting socially shared regulation of learning in CSCL: Progress of socially shared regulation among high- and low-performing groups. *Computers in Human Behavior*, *52*, 562–572. https://doi.org/10.1016/j.chb.2015.03.082.

Mor, Y., & Craft, B. (2012). Learning design: Reflections upon the current landscape. *Research in Learning Technology*. https://doi.org/10.3402/rlt.v20i0.19196

Peffers, K., Tuunanen, T., Rothenberger, M., & Chatterjee, S. (2007). A design science research methodology for information systems research. *Journal of Management Information Systems*, *24*, 45–77.

Rodríguez-Triana, M. J., Martínez-Monés, A., Asensio-Pérez, J. I., & Dimitriadis, Y. (2015). Scripting and monitoring meet each other: Aligning learning analytics and learning design to support teachers in orchestrating CSCL situations. *British Journal of Educational Technology*. https://doi.org/10.1111/bjet.12198

Saint, J, Gasevic, D., Matcha, W., Ahmad U., N., & Pardo, A. (2020). Combining analytic methods to unlock sequential and temporal patterns of self-regulated learning. *Companion Proceedings 10th International of Conference on Learning Analytics & Knowledge (LAK 2020)*, 402–411.

Saint, John, Whitelock-Wainwright, A., Gasevic, D., & Pardo, A. (2020). Trace-SRL: A Framework for Analysis of Micro-Level Processes of Self-Regulated Learning from Trace Data. *IEEE Transactions on Learning Technologies*. https://doi.org/10.1109/TLT.2020.3027496

Siadaty, M., Gasevic, D., & Hatala, M. (2016). Trace-Based Microanalytic Measurement of Self-Regulated Learning Processes. *Journal of Learning Analytics*, *3*.

Voogt, J., & Roblin, N. P. (2010). 21st century skills. *Discussienota. Zoetermeer: The Netherlands: Kennisnet*, *23*(03), 2000.