Computers & Education 136 (2019) 34–48

Contents lists available at ScienceDirect

Computers & Education

journal homepage: www.elsevier.com/locate/compedu

Examining sequential patterns of self- and socially shared regulation of STEM learning in a CSCL environment

Juan Zheng[a], Wanli Xing[b,*], Gaoxia Zhu[c]

[a] Department of Educational Counselling & Psychology, McGill University, Montreal, Quebec, Canada
[b] Department of Educational Psychology & Leadership, Texas Tech University, Lubbock, TX, 79409, USA
[c] Ontario Institute for Studies in Education, University of Toronto, Toronto, ON, M5S 1V6, Canada

A R T I C L E I N F O

Keywords:
Self-regulated learning
Socially shared regulation
STEM education
Sequential mining

A B S T R A C T

While the importance in collaborative learning of both self- and socially shared regulatory processes has increasingly been emphasized, little research has examined sequences of such processes and how they influence group performance. This study identified sequences of self- and socially shared regulatory activities in the online chats and logs of students completing a STEM task in a computer-supported collaborative learning (CSCL) environment. High school and college students (N = 156) were randomly assigned to groups of three and asked to solve four tasks of increasing complexity in a virtual learning environment designed to teach students about electronics. The results revealed that the students engaged mostly in executing, self-monitoring, and socially shared monitoring activities. The successful groups demonstrated more frequent and more diverse regulatory activities than did the less successful groups. A Markov chain analysis revealed that the successful groups were most likely to start with self-executing and end with socially shared monitoring, while the less successful group were most likely to start with executing and end with self-executing. The results of this study reveal that the timing of socially shared monitoring influences the success of collaborative learning, which have implications for teaching practices and for adaptive scaffolding group learners in CSCL.

1. Introduction

With the technological developments of the past two decades, extensive research has been conducted into computer-supported collaborative learning (CSCL). This research has found that CSCL technologies can improve both individual and collective learning outcomes by improving group interactions and promoting the construction of mutual knowledge (Järvelä & Hadwin, 2013). Effective interactions can result in efficient, successful, and enjoyable learning, while interactions in which effort is not shared can cause group members to be less efficient and less successful. For this reason, it is important to determine why some collaborative work is less successful and how to better facilitate successful collaborative work.

A growing body of research has examined regulation mechanisms used in CSCL to determine how groups and individuals engage in, sustain, support, and productively complete processes of collaborative learning. Self-regulated learning (SRL) and the socially shared regulation of learning (SSRL) are two concepts describing these processes of collaborative learning (Hadwin, Järvelä, & Miller, 2011). When they engage in SRL, individuals purposefully adjust their cognitions, behaviors, motivations, and emotions to achieve their learning goals (Zimmerman, 2000). In the meanwhile, groups engaged in SSRL use various social processes to regulate their

* Corresponding author.
E-mail addresses: juan.zheng@mail.mcgill.ca (J. Zheng), wanli.xing@ttu.edu (W. Xing), gaoxia.zhu@mail.utoronto.ca (G. Zhu).

https://doi.org/10.1016/j.compedu.2019.03.005
Received 3 November 2018; Received in revised form 12 March 2019; Accepted 13 March 2019
Available online 16 March 2019
0360-1315/ Published by Elsevier Ltd.

joint work (Volet, Vauras, & Salonen, 2009). While SRL and SSRL are seemingly two strands of research—SRL demonstrates the role of regulation in individualized learning environments, while SSRL highlights interactions among group members engaged in collaborative tasks, however, they share common characteristics. Both are triggered by authentic challenges and display a cyclical nature, making occurrences and sequences of regulatory activities dynamic (Järvelä, Malmberg, & Koivuniemi, 2016; Zimmerman, 2008). For example, new information, which can be introduced by group members and come from external resources, is one type of challenge that can trigger changes in SRL to SSRL (Winne, 2015). The dynamic occurrences of regulator activities enable examinations of regulatory patterns at the individual and group levels, as well as the congruence between regulation at the individual level (via SRL) and regulation at the group level (via SSRL).

Recent studies have separately described transitions between regulatory activities, focusing either on SRL patterns—which have been identified using think-aloud protocols (Bannert, Reimann, & Sonnenberg, 2014)—or on socially shared regulatory patterns—which have been identified in discussion messages (Kwon, Liu, & Johnson, 2014). As a result, little research has investigated how SRL and SSRL activities co-occur in authentic tasks and how different regulatory patterns influence group performance. However, new technologies for promoting learning interaction have provided new sources of trace data (e.g., log files and discussion messages) with which researchers can observe transitions between self- and group-level regulatory activities. These technologies have also yielded new analytical techniques that researchers can use to identify patterns in such activities (Reimann, Markauskaite, & Bannert, 2014). This study uses some of these tools to investigate the temporal and sequential characteristics of SRL and SSRL regulation (i.e., SRL and SSRL) in computer-supported collaborative learning, aiming at informing teachers' facilitation of group learning to yield better performance.

2. Theoretical framework

SRL and SSRL are two important types of regulated learning employed in collaborative learning environments. Regulated learning extends the level of learning from simple knowledge construction to cognitive, motivational, emotional, behavioral, and metacognitive processes (Järvelä & Hadwin, 2013). Individuals in groups simultaneously regulate their own learning processes (via SRL) and collectively regulate group processes in a synchronized and productive way (SSRL).

2.1. Self-regulated learning

SRL is an overarching construct that describes active and goal-directed processes by which learners adjust their own cognition, motivation, emotion, metacognition, and behavior (Pintrich, 2000; Winne & Perry, 2000; Zimmerman, 2000). There is a general consensus among SRL researchers that SRL occurs in a number of cyclical phases (Zimmerman, 2008). Winne and Hadwin (1998) propose a cyclical model that incorporates four phases: “task understanding,” “goal setting and planning,” “strategic enactment,” and “strategic adaptations.” The current study uses this model because these four phases are related to the context of solving STEM tasks. In solving STEM tasks by applying Ohm's Law and exploring the relationship between resistance and voltage, the learners in this study needed, in the first phase (task understanding), to have perceptions of what was asked of them, e.g., to obtain a specific voltage for a given resistance. In the second phase (goal setting and planning), the learners needed to think about the actions they needed to take to solve the task. For example, the learners could set goals for adjusting the resistance, changing circuits, and using the digital multimeter (DMM) to measure the voltage, current, or resistance. The third phase (strategic enactment) includes the execution of actions planned in the second phase as well as elaboration, which includes the use of deep thinking and reasoning to confirm or disconfirm hypotheses. In elaboration, for example, the students could use a formula to calculate the specific resistor value needed to accomplish the task. In the fourth phase (strategic adaptations), learners could change and update strategies in response to challenges throughout the task, during which monitoring has been highlighted as essential for the effectiveness of strategic adaptation (Malmberg, Järvelä, & Järvenoja, 2017; Winne & Hadwin, 1998). Monitoring represents students' ability to keep track of their own progress and evaluate the time and effort spent in the task. However, not all learners are capable of accurately employing monitoring processes (Winne & Jamieson-Noel, 2002), making monitoring a crucial process in determining the strength of SRL. Winne and Hadwin's (1998) model describes four phases of SRL and illustrates the dynamic nature of SRL. The conceptualization of these four phases of SRL informs the operation and definition of SRL in this current study.

Referencing to Winne and Hadwin's (1998) SRL model and taking account of the specificity of STEM tasks, we focused on examining how students analyze their task in the first phase, make a plan for themselves in the second phase, elaborate their thinking in the third phase, and monitor their progress in the fourth phase. Task analysis, planning, elaboration, and monitoring reveal how students regulate themselves to solve a STEM task. These four critical SRL processes could also advance our understanding of SRL in STEM learning by unraveling the characteristics and patterns of SRL in collaborative settings.

Guided by the challenges of STEM tasks, learners take responsibility for their thinking processes by analyzing tasks, making plans, monitoring and elaborating their learning progress (Zimmerman, 2000). These self-regulatory processes evolve via a recursive cycle in which, at any stage, learners refine their perceptions of their tasks or their goals (McCardle & Hadwin, 2015). However, SRL alone is insufficient to guarantee successful collaborative learning. When learners self-regulate during collaboration, they are individually responsible for their contributions to the group. More importantly, they can even externalize their SRL, contributing to socially shared regulation in collaborative interactions (Järvelä et al., 2016).

2.2. Socially shared regulation of learning

SSRL is a group-level phenomenon in which students collectively share and negotiate cognitive, behavioral, motivational, and emotional conditions to reach a common perception of a collaborative task (Isohätälä, Järvenoja, & Järvelä, 2017; Winne, Hadwin, & Perry, 2013). Panadero and Jarvela (2015) conducted a qualitative review to analyze empirical evidence for socially shared regulation, and they affirmed the existence of SSRL, even though it had been conflated with socially shared metacognition. SSRL differs in several respects from other types of regulation (including SRL and co-regulation), including who engages in regulation within the group and what the group regulates (Rogat & Linnenbrink-Garcia, 2011). First, SSRL activities include multiple group members, albeit with potentially different levels of involvement. SSRL situations can range from one student temporarily playing a more instructive role to multiple group members equally and jointly regulating an activity (Vauras, Iiskala, Kajamies, Kinnunen, & Lehtinen, 2003). Second, SSRL involves different aspects and levels of regulation to ensure that group members remain engaged and offer consistent efforts. Low-level SSRL can involve simply exchanging information or clarifying a misunderstanding (Rogat & Linnenbrink-Garcia, 2011). In contrast, high-level SSRL is characterized by deep content processing, during which multiple students elaborate on each other's ideas and jointly monitor contributions (Rogat & Linnenbrink-Garcia, 2011; Volet, Summers, & Thurman, 2009). The level of SSRL achieved can be influenced by who is doing the regulating, what is being regulated, and when it is being regulated.

SSRL occurs when group members have complementary perceptions of their task, collectively set common goals, share responsibility for strategic enactment, and coordinate changes and adaptations to optimize collaboration across tasks (Miller, Järvelä, & Hadwin, 2017). Group members achieve collaborative awareness of a task by collectively adjusting their metacognition, cognition, motivation, and behavior (Hadwin et al., 2011). SSRL does not necessarily require sameness among group members in terms of thoughts but instead emphasizes the processes of negotiation towards sameness when each group member actively contributes to a joint perception of the task or to strategies for achieving the common goal (Järvelä & Hadwin, 2013).

Järvelä et al. (2016) concretized the negotiation processes into SSRL task understanding, SSRL planning, and SSRL strategy use. Malmberg et al. (2017) further highlight the crucial role of SSRL monitoring in addressing group challenges to reach a consensus. Accordingly, the current study applied these SSRL processes into the context of STEM learning. In align with the aforementioned SRL model of Winne and Hadwin (1998), we identify SSRL task analysis, SSRL monitoring, SSRL elaboration, and SSRL monitoring are the key group-level regulatory processes to influence and determine the effectiveness of collaborative learning.

To solve STEM problems in a CSCL context, students need to discuss the tasks to achieve common understanding (i.e. engage in "task analysis"), set up plans for completing the tasks (i.e. engage in "planning"), verbalize and justify their strategies (i.e. engage in "elaboration"), and manage time and conflicts to maintain common understanding (i.e. engage in "monitoring"). As are those of SRL, the processes of SSRL are dynamic and recursive: group members may collectively revisit task perceptions or reset common goals in response to ever-changing challenges. These challenges can include new information from external resources or unhealthy group interactions. For example, when group members do not interact effectively or in a friendly manner, planning and monitoring at the group level may be needed to promote high-quality regulatory processes and to ensure group understanding (Rogat & Linnenbrink-Garcia, 2011). This also indicates that the timing of SSRL matters as early group-level regulation can facilitate the establishment of common ground and thereby reduce the commitment required for maintaining group effort (Lajoie & Lu, 2012).

2.3. Sequential characteristics reveal the interplay between SRL and SSRL

Researchers have widely recognized the central role of SRL and SSRL in collaborative learning. Individuals in a group must regulate themselves while regulating collaboratively to guide and support the construction of group knowledge (Jarvela, Volet, & Jarvenoja, 2010). Self-and socially shared regulation, which are important parts of successful collaboration, arise simultaneously and reciprocally influence each other (Hadwin et al., 2011). For example, individuals in groups negotiate common goals, and everyone adjusts their own goals in accordance with their personal competency. Conversely, individuals who feel they are not able to achieve their personal goals on time may trigger discussions aimed at resetting common goals or monitoring group progress. According to Järvelä and Hadwin (2013), SRL sets the stage for SSRL, and SSRL can mediate SRL processes (e.g., planning and monitoring) by changing the conditions of the collaborative environment. Other forces may also drive shifts between SRL and SSRL, including new information introduced by group members (including group facilitators) and from external resources (including technology-supported feedback) (Winne, 2015). For these reasons, researchers should consider the interplay between SRL and SSRL processes as they unfold in authentic learning contexts. Mapping how regulation is manifested by individuals and groups and how regulatory activities oscillate between SRL and SSRL is an important line of inquiry that deserves attention (Winne, 2015).

Advances in identifying and tracking regulatory activities—e.g., think-aloud, text mining conversational messages, and analyzing log files—in real time have improved the diversity of analyses used to reveal the dynamic attributes of SRL and SSRL, including frequency, duration, and time-dependent patterns (Azevedo, 2014). These methods offer new opportunities to examine the temporal and sequential characteristics of regulatory activities. Examining both SRL and SSRL, Malmberg et al. (2017) analyzed videos of conversations that students engaged in during face-to-face activities in regulated collaborative learning. They found that self-level task execution promoted the occurrence of socially shared planning and that monitoring in both SRL and SSRL facilitated progress in executing tasks. Splichal, Oshima, and Oshima (2018) examined internal scripts generated by students in a CSCL environment and found that social–emotional challenges (e.g., being overruled or undermined in social interactions) contributed to self-regulation and that social–cognitive challenges (e.g., knowledge-related discussion and negotiation) triggered socially shared regulation. However, few research has examined the sequential patterns of both SRL and SSRL using chats and log files in a CSCL environment.

Online chats and log files make it possible to examine regulatory processes at both individual level and group level. For example, Jarvela et al. (2016) identified self-regulatory activities from log file traces of individual student's activities and assesses group activities at the social level from the online chats in collaborative learning space. Malmberg et al. (2017) relied on communication within group members to identify regulatory activities at both the individual level and group level. They differentiated SRL and SSRL based on who was regulating and what was being done. These prior studies set good examples for simultaneously examining both SRL and SSRL in CSCL.

The current study synthesized the four SRL processes (i.e., SRL task analysis, SRL planning, SRL elaboration, SRL monitoring) and four SSRL processes (i.e., SSRL task analysis, SSRL planning, SSRL elaboration, SSRL monitoring) with the intention of adapting them into STEM learning context. In addition, how students proceeded the task (i.e., execution) was included in the current study to better inform how students switch between SRL and SSRL. It is possible that task execution, SRL activities, and SSRL activities could influence how the group performed in a STEM task. Following the aforementioned methodology in identifying SRL and SSRL processes from online chats and log files, this study aims at examining the characteristics of both SRL and SSRL and how SRL and SSRL relate to group performance. We sought to answer three research questions:

(1) How do self- and socially shared regulation activities occur in groups that successfully and less successfully complete group tasks?
(2) Does the variation of regulatory activities differ between groups that successfully and less successfully complete the tasks?
(3) Are there differential regulatory patterns associated with the performance of the groups in solving tasks?

3. Methods

3.1. Participants and settings

The participants in this study were 144 students from 5 colleges and 12 high schools in the United States. At each college and high school, the students in each class were grouped randomly into threes and asked to collaboratively solve four tasks of increasing complexity. These tasks were related to electronics and specifically to Ohm's Law. Each student was given control of a resistor in a series circuit and then asked to figure out the resistance value for the resistor they controlled. This required working with the other members of their group. However, the students were assigned to groups of three by their teachers, and they had no idea of who their group members were. Group members sat around the computer lab and could only communicate using the chat box of the simulation platform. As the tasks increased in complexity, the students also needed to collaboratively calculate the supply-voltage value and the value for the external resistor.

Textual instructions and videos were embedded in the platform to help the participants to understand the tasks and the platform. The teacher was not involved in the process of completing the tasks. The students were given about 60 min to work on the four tasks and were allowed to work on the tasks in any order. They worked in a simulated online collaborative inquiry learning environment called Teaching Teamwork, which included a database of interactive STEM activities. As shown in Fig. 1 (a), the platform interface was divided into four areas: a) the upper-left area showed the goal for each task, b) the middle-left area showed the resistor that the student controlled, c) the bottom-left area showed the series circuit, and d) the right area was the chat window. There were also two

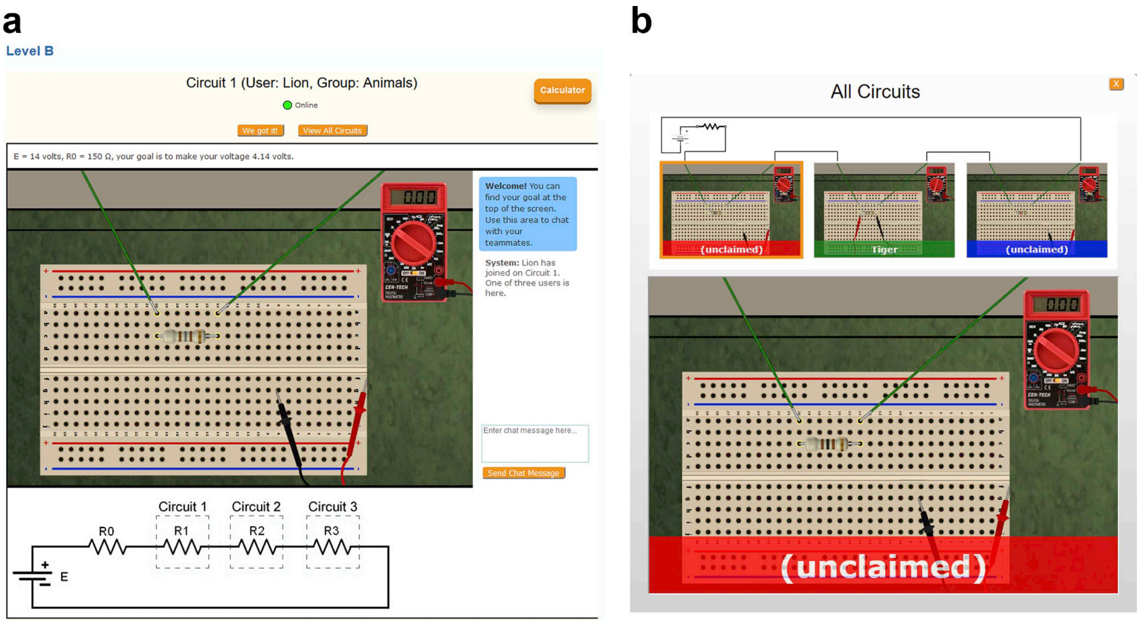


Fig. 1. (a). The interface of a Level B simulation task. (b). The interface of the zoom-in function.

tools available to the participants: a) a digital multimeter to help them measure the resistance of the resistor or the current and voltage across the resistor and b) a calculator. As shown in Fig. 1(b), the group members can also zoom in (view all circuits) to see the resistance of all the three resistors in a series circuit. They can measure the resistance, current, and voltage of each resistor, which may help them track the status of their group problem-solving and, therefore, help them adjust their plan and actions. All of the chat messages and each participant's behavior in the platform—including changing the resistance, inputting numbers into the calculator, and measuring results— were recorded in the log files.

3.2. Analysis

We analyzed the students' behavioral data and the history of chats they generated while completing each task. The tasks that were solved successfully were labeled as successful condition, while the tasks that were not solved were considered as the less-successful condition. The analysis can be divided into three layers. To respond to the first research question, we classified the behavioral-log data and chat messages of the students in the "successful" and "less successful" conditions into the subcategories of "self-regulation" and "socially shared regulation" using a supervised machine learning model. To answer the second research question, we conducted an entropy analysis in which each subcategory of regulatory activities was manipulated to generate the entropy value for groups in both "successful" and "less successful" condition. To answer the third research question, we used Markov chain analysis and sequential pattern mining to identify sequences in the two conditions. We describe the process and the instruments in detail in the following subsections.

3.2.1. Coding framework for developing training data

The online chats and behavioral log files were coded and categorized according to the coding framework in Table 1, which was developed based on the SRL model (Winne & Hadwin, 1998) and SSRL theoretical foundation discussed before. Specifically, behavioral-log data could directly reflect some aspects of self-regulation (i.e., SRL elaboration, SRL monitoring) and execution. SRL elaboration was indicated by strategic actions students utilized, namely "perform calculation" to achieve their learning goals. SRL monitoring referred to students' actions in checking their personal progress. By "Opening Zoom View" and "Selecting Circuit in Zoom", students could keep track of their own progress and make adaptations accordingly. Execution is the regular individual actions each student performed to achieve group goals, namely "change circuit" and "measure with a digital multimeter (DMM)". Although SRL elaboration, SRL monitoring, and execution could be directly calculated from behavioral-log data, the behaviors in the log data did not reveal how students understand and plan for the task. SRL task analysis and SRL planning needed to be derived from the chat messages where students exchanged information with group members. Therefore, two SRL activities and four SSRL activities were coded from chat messages. As discussed earlier, SRL and SSRL could be differentiated by who was regulating and what was being done (Malmberg et al., 2017). If students were activating their personal knowledge and thinking about personal actions, they were engaged in SRL task analysis and SRL planning. Self-directed words, such as "I am", "I need" and "Let me" were evident features of regulatory activities at the individual level. In contrast, SSRL task analysis and SSRL planning are manifested by statements which centered on group tasks and group actions. SSRL elaborating are the statements where students were making transparent reasoning to the whole group, while SSRL monitoring are instructions guided the whole group towards managing time, solving conflicts, and attaining a mutual understanding. Taken together, online chats and behavioral log files can be used to simultaneously and mutual-exclusively extract both SRL and SSRL activities.

3.2.2. Detection of self-regulation and socially shared regulation

Given that the groups produced thousands of lines of chat, it would have been quite difficult to manually analyze their messages, as have many small-sample CSCL studies (Kwon et al., 2014). Many previous studies have shown that it is possible to build machine learning models that can automatically identify theoretical constructs in conversations (Mu, Stegmann, Mayfield, Rose, & Fischer, 2012; Zhu et al., 2019; Xing and Gao, 2018; Xing et al., 2018; Popov et al., 2018). This process usually involves three steps. In the first step, human coders manually determine what kinds of constructs chat messages contain and then build a training dataset. In the second step, various features are extracted from the training data and a text-classification model is built using various supervised machine learning models. The performance of this model is then evaluated. In the last step, the rest of the chats are entered into the supervised machine learning model.

In this study, we used a similar approach to develop a semi-automatic workflow (see Fig. 2) that could recognize self- and social-level regulation constructs in the chat messages. First, two senior researchers randomly selected and coded the chat history of 30 solved and unsolved tasks after developing a shared understanding of the framework presented in 3.2.1. In total, 886 pieces of chat message were coded for the SRL task analysis and SRL planning as well as SSRL ask analysis, SSRL planning, SSRL elaborating, SSRL monitoring and execution. The inter reliability between the two researchers of the coding was very good (Kappa = 0.844), which was calculated based on 308 chat messages coded by both researchers. Disagreements between these two raters were discussed and resolved. The results were treated as training data and served as the ground truth in evaluating the performance of the model. Second, the features of the text were extracted mainly based on three techniques: linguistic feature (e.g., "the length of sentence", "sentence type," and "part of speech", etc.), LIWC-related features (e.g., "pronoun", "tense", "positive or negative sentiment", etc.) (Pennebaker, Boyd, Jordan, & Blackburn, 2015) and Latent Dirichlet allocation -topical features (the number of different topics in the text). The comments from students in the course forum may relate to different aspects of regulation, including discussing the task, greeting, communicating, etc. For a more detailed discussion of the features, please refer to Wang, Kraut, and Levine (2012, February). In order to classify each comment, four machine learning classification algorithms (decision tree, naïve Bayes, support vector

Table 1
The coding framework for self-regulation and socially shared regulation.

Category	Description	Example	Data Source	Features
SRL-Task Analysis	Activating previous knowledge related to the task and its contents; thinking about the purpose of the task;	ok i need 6.69; I am the bear;	Chat messages	I am; I need
SRL-Planning	Thinking about the personal actions needed to the solve the task;	let me get 1v hold up; let me readjust;	Chat messages	Let me
SRL-Elaborating	Connecting to the goal of the task by reasoning. Using the formula to calculate the specific value of the resistor needed to achieve the goal.	"Calculation performed"	Log data-Calculation performed	
SRL-Monitoring	Trying to create an overview of one's personal progress and the current status of the whole group.	"Opened Zoom View"; "Selected Circuit in Zoom"	Log data-Check all the circuits	
SSRL-Task analysis	Talking about the group task and trying to figure out the collaborative setting	how many volts do you guys need? we need the same values probably	Chat messages	Numbers, e.g. 320, 640; Goal
SSRL-Planning	Forming a statement or set of statements directed towards the other group members concerning the next actions that should be taked to complete the task.	everyone set to 180 - see what that does;	Chat messages	let's; go up; down; drop; we need
SSRL- Elaborating	Elaborating one's thoughts to make one's reasoning transparent to the group	so there is a 3.76 drop across r = 560 totals to 16	Chat messages	
SSRL-Monitoring	Managing one's time, giving group instructions, engaging in dialogue to resolve conflict, and attaining mutual understanding	Who opened the circuit? Let's get move on;	Chat messages	Who; why; tell me
Execution	Testing one's own ideas or following the instructions of group members to adjust the resistors by making them lower or higher	"change circuit" and "DMM measurement"	Log data-Change circuit and measure DMM	

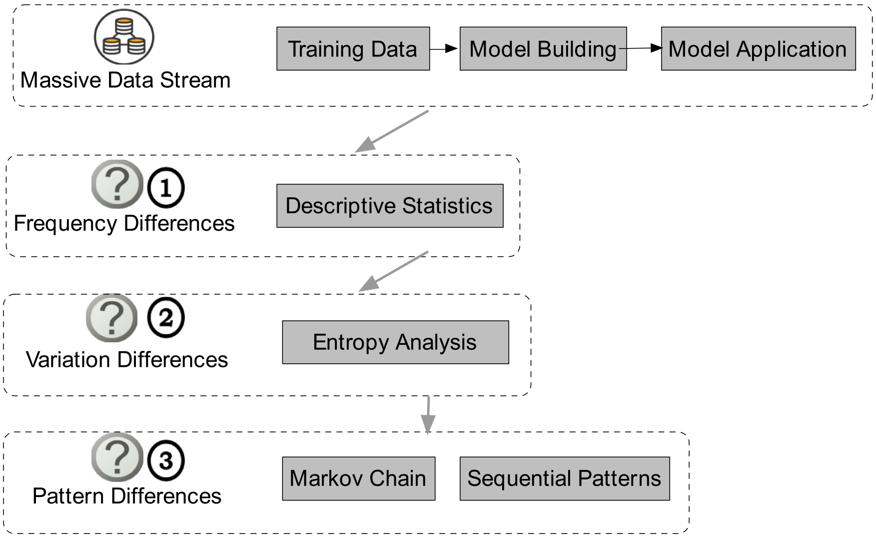


Fig. 2. Methodological workflow.

machines, and random forest) were trained based (Kotsiantis, Zaharakis, & Pintelas, 2007). A 10-fold cross-validation method was applied to evaluate the performances of the different classification models. For a simple demonstration of the feature extraction and classification, please refer to Appendix 1. Finally, we applied the model with the best generalization ability and performance to the rest of the unlabeled chat messages to automatically detect self- and social-level regulation.

3.2.3. *Entropy analysis*

Entropy analysis has been used in social studies to measure the participatory diversity of students' online communications in CSCL environments (Matei, Oh, & Brun, 2006), to understand the relationship between small-group participation and learning performance (Bruno, 2010), and to evaluate the depths of understanding of students in different communities (Xing, 2017). In this study, for each group process in the successful and less successful condition, we calculated its entropy value in order to reflect how diverse and stable the regulation state changes over time. The entropy index was calculated using the Shannon entropy formula:

$$H(i) = -\sum_{i=1}^{9} p(i)\log_2 p(i).$$

where *p(i)* is the probability of a regulation state *i* appearing in a given stream of regulated activities and 9 represents the total number of shared regulation states. The minimum entropy value is 0, which represents no changes in the regulation state over the collaboration process, while the maximum value is 1, which represents the maximum number of changes and degree of diversity across all of the regulation states.

3.2.4. *Markov chain analysis and sequential pattern analysis*

To identify instances of the nine regulatory states, we conducted Markov Chain analysis and sequential pattern mining. A Markov Chain is a stochastic model that describes a sequence of possible states. Several studies in CSCL have used Markov models to explain sequences in collaboration (Soller, 2004). The result of a Markov Chain analysis is a matrix that explains the probabilities of transitions between different states. Unlike Markov Chain analysis, sequential pattern mining identifies in a set of subsequences that occur above a set frequency threshold. The length of a subsequence can be more than two. Table 2 provides a brief example of the results of sequential pattern mining. The sequence (EX) → (EX) is a subsequence of Sequence 2, 3, and 4. The probability that subsequence (EX) → (EX) will occur, or its "support value" is 3/4 (0.75). The sequence (SSRL-TA) → (SSRL-TA) → (SSRL-PL) → (SSRL-EL) only appears in Sequence 1, so its support value is 0.25. The results of sequential pattern mining are different subsequences that meet the

Table 2
An example dataset of sequences obtained through sequential pattern mining.

Sequence ID	Sequence
1	(SSRL-TA) → (SSRL-TA) → (SSRL-PL) → (SE) → (SE)
2	(SRL-EL) → (SSRL-MO) → (EX) → (EX)
3	(SSRL-MO) → (EX) → (EX) → (SRL-MO)
4	(EX) → (EX) → (EX) → (SSRL-TA) → (SSRL-MO)

Table 3
Performance of the text-classification models.

Features	Decision Tree		Naïve Bayes		SVM		Random Forest	
	Precision	Recall	Precision	Recall	Precision	Recall	Precision	Recall
LF	0.77	0.47	0.68	0.39	0.37	0.36	0.71	0.31
LIWC	0.69	0.26	0.65	0.26	0.66	0.26	0.66	0.23
Topic Features	0.68	0.29	0.64	0.25	0.63	0.23	0.68	0.30
LF, LIWC	0.80	0.50	0.46	0.41	0.75	0.57	0.71	0.27
LF, Topic	0.75	0.33	0.66	0.31	0.69	0.34	0.67	0.26
LIWC, Topic	0.77	0.45	0.67	0.37	0.74	0.49	0.73	0.34
LF, LIWC, Topic	0.84	0.89	0.48	0.53	0.65	0.74	0.84	0.88

*LF: Linguistic Features; *Topic: Topical Features.

threshold. If we set 0.7 as the threshold, for instance, (EX) → (EX) will be picked up, but not (SSRL-TA) → (SSRL-TA) → (SSRL-PL) → (SSRL-EL).

4. Results

4.1. How do self- and socially shared regulation activities occur in groups that successfully and less successfully complete group tasks?

Using various algorithms, including supervised machine learning algorithms, text-classification models were built that could semi-automatically identify instances of self- and socially shared regulation in the chat messages. The performance of each classification model was evaluated using precision and recall measures, as shown in Table 2. The results revealed that the decision tree algorithm had the best performance when the linguistic features, LIWC features, and topical features were all used. The precision of the decision tree was as high as 84.0%, and its recall rate was 88.6%. Based on a review of text-classification models, this quality of performance was considered "good to excellent" (Dalal & Zaveri, 2011). We then applied this decision model to the rest of the chat messages to automatically detect instances of self-regulation (task analysis and planning) and socially shared regulation (socially shared task analysis, socially shared planning, socially shared elaborating, and socially shared monitoring). For self-regulation, the values for "executing" and "elaborating" were directly computed from the log data. The statistics for each dimension of regulation are shown in Table 3 in the next section.

The groups produced a total of 31,230 regulatory events. A visualization of all of these regulatory activities (see Fig. 3) reveals that task execution, SRL monitoring, and SSRL monitoring were the most frequent activities in both the "successful" and "less successful" conditions. In contrast, SRL planning occurred the least frequently (see Table 4). Of the SRL activities, the students engaged most in SRL monitoring (f = 4946) and engaged least in SRL planning (f = 18). Similarly, of SSRL activities, the students devoted the most effort to SSRL monitoring (f = 4184) and the least effort to SSRL planning (f = 102). The students in the "successful" condition (M = 240.04, SD = 420.14) demonstrated more regulatory activities than did the students in the "less successful" condition (M = 182.12, SD = 178.01). The students in the "successful" condition engaged more frequently in SRL monitoring and SRL elaborating and less frequently in SRL task analysis and SRL planning than did the students in the "less successful" condition. In contrast, compared to the students in the "less successful" condition, the students in the "successful" condition demonstrated more

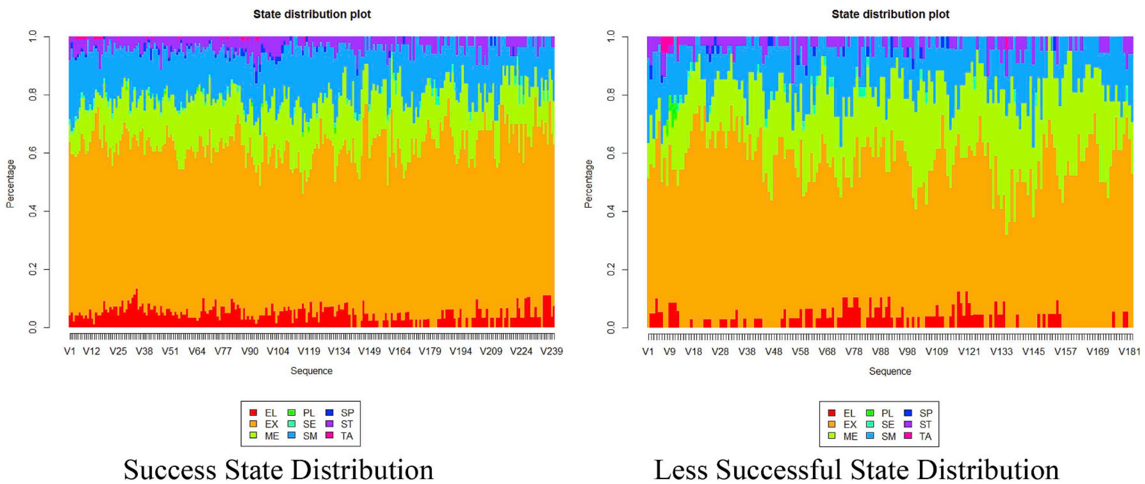


Fig. 3. Overall proportion of regulatory activities in the "successful" and "less successful" conditions. * TA: Task analysis; PL: Planning; EL: Elaboration; MO: monitoring; EX: Execution.

Table 4
Statistics for the "Successful" and "Less Successful" conditions.

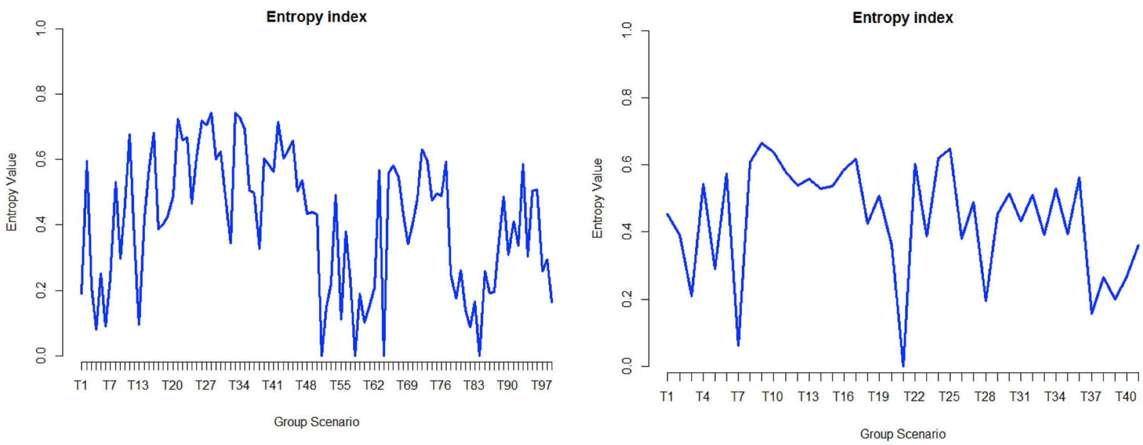
Dependent Variable	Performance	Means	SD
SRL Task Analysis	1	0.24	2.41
	0	0.27	0.55
SRL Planning	1	0.09	0.73
	0	0.22	0.47
SRL Elaborating	1	8.87	13.44
	0	4.73	6.23
SRL Monitoring	1	34.07	75.59
	0	38.37	53.36
SSRL Task analysis	1	7.45	9.57
	0	5.83	8.86
SSRL Planning	1	0.78	2.30
	0	0.61	1.67
SSRL Elaborating	1	1.03	3.40
	0	0.98	2.54
SSRL Monitoring	1	29.78	28.60
	0	30.15	45.82
Executing	1	157.73	406.18
	0	100.98	126.08

* < 0.05, ** < 0.01, *** < 0.001.
* 1: task solved, 0: task unsolved.

SSRL task analysis, SSRL planning, and SSRL elaboration, but less SSRL monitoring. Task execution also occurred more frequently in the "successful" condition than it did in the "less successful" condition.

4.2. Does the variation of regulation activities differ between groups that successfully and less successfully complete the tasks?

Entropy values were calculated to represent the variation among the regulatory activities that a group engaged in while completing a task. Lower entropy values indicated consistency in students' regulatory activities, while higher entropy values indicated instability in students' regulatory activities. The entropy values for each task completed by the groups in both the "successful" and "less successful" conditions are visualized in Fig. 4. As Fig. 4 shows, the entropy values for the groups in the "successful" condition fluctuated between 0 and 0.74. The entropy values for the groups in the "less successful" condition ranged from 0 to 0.67. A Welch two-sample *t*-test was performed to compare the entropy values of the groups in the "successful" and "less successful" conditions. The results revealed a significant difference between the entropy values for the "successful" and "less successful" conditions ($t = -2.86$, $df = 77$, $p = 0.0054 < 0.01$). The regulatory activities of the groups in the "successful" condition ($M = 0.41$, $SD = 0.2$) had less variability than did the regulatory activities of the groups in the "less successful" condition ($M = 0.44$, $SD = 0.17$), indicating that the groups in the "successful" condition engaged in more consistent regulatory activities. In other words, the regulatory activities of the groups in the "successful" condition were relatively stable, while the regulatory activities of the groups in the "less successful" condition were likely to change periodically as they completed a task.



Successful condition

Less successful condition

Fig. 4. The entropy value of for each individual in the "successful" and "less successful" conditions.

	SRL-TA	SRL-PL	SRL-EL	SRL-MO	SSRL-TA	SSRL-PL	SSRL-EL	SSRL-MO	EX
SRL-TA	0.35	0.00	0.00	0.00	0.00	0.01	0.00	0.00	0.00
SRL-PL	0.04	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00
SRL-EL	0.00	0.00	0.36	0.04	0.04	0.08	0.05	0.07	0.01
SRL-MO	0.22	0.00	0.14	0.69	0.08	0.20	0.08	0.08	0.04
SSRL-TA	0.00	0.00	0.05	0.02	0.27	0.04	0.02	0.07	0.01
SSRL-PL	0.09	0.28	0.01	0.00	0.00	0.08	0.00	0.01	0.00
SSRL-EL	0.00	0.00	0.01	0.00	0.00	0.00	0.15	0.01	0.00
SSRL-MO	0.05	0.14	0.20	0.07	0.31	0.25	0.36	0.45	0.06
EX	0.26	0.57	0.25	0.17	0.30	0.33	0.33	0.32	0.87

Fig. 5. Heat map of the transitions between regulatory activities of the students in the "successful" condition.
* TA: Task analysis; PL: Planning; EL: Elaboration; MO: monitoring; EX: Execution.

4.3. Are there differential regulatory patterns associated with the performance of the groups in solving tasks?

A Markov Chain analysis was performed to determine the transition probabilities in the "successful" condition and the "less successful" condition regarding self- and socially shared regulatory activities. The transition model fits indices in the "successful" condition were: LogLikelihood = −17589.38, AIC = 35218.76, BIC = 35380.27. The groups in the "successful" condition were most likely to start with task execution ($p = 0.60$) and to end with SSRL monitoring ($p = 0.51$). The transition matrix for the regulatory activities of the groups in the "successful" condition is displayed inFig. 5. The results show that the groups in the "successful" condition were most likely to get immersed in execution ($p = 0.87$) and SRL monitoring ($p = 0.69$). The probability of a transition between from execution to SRL planning is 0.57. Of the self-regulatory activities, the students had the highest probability of transitioning from SRL monitoring to SRL task analysis ($p = 0.22$). Of the socially shared regulatory activities, the groups had the highest probability of transitioning from SSRL monitoring to SSRL task analysis ($p = 0.31$), SSRL planning ($p = 0.25$), and SSRL elaboration ($p = 0.36$). These results reveal the impacts of both SRL monitoring and SSRL monitoring on driving other regulatory activities. Of the transitions between SRL activities and SSRL activities, SSRL planning was likely to drive students toward SRL planning ($p = 0.28$), and SSRL monitoring was likely to drive students towards SRL task analysis ($p = 0.22$).

The transition matrix of the regulatory activities of the students in the "less successful" condition is displayed in Fig. 6. The transition model fit indices in the less successful condition were: LogLikelihood = −6823.16, AIC = 13686.32, BIC = 13824.69. The groups in the "less successful" condition were most likely to start with task execution ($p = 0.51$) and end with task execution ($p = 0.41$). Like the students in the "successful" condition, the students in the "less successful" condition were also absorbed in execution ($p = 0.66$) and SSRL monitoring ($p = 0.79$). They demonstrated a high probability of transitioning from execution to SRL

	SRL-TA	SRL-PL	SRL-EL	SRL-MO	SSRL-TA	SSRL-PL	SSRL-EL	SSRL-MO	EX
SRL-TA	0.00	0.00	0.01	0.00	0.00	0.00	0.00	0.00	0.00
SRL-PL	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00
SRL-EL	0.00	0.11	0.30	0.02	0.01	0.04	0.05	0.03	0.12
SRL-MO	0.09	0.44	0.17	0.66	0.17	0.04	0.03	0.12	0.08
SSRL-TA	0.00	0.11	0.02	0.02	0.16	0.43	0.03	0.06	0.02
SSRL-PL	0.00	0.00	0.01	0.00	0.00	0.17	0.00	0.01	0.00
SSRL-EL	0.00	0.00	0.01	0.00	0.00	0.00	0.05	0.17	0.00
SSRL-MO	0.27	0.00	0.16	0.10	0.36	0.30	0.43	0.43	0.10
EX	0.64	0.33	0.33	0.19	0.29	0.39	0.43	0.34	0.79

Fig. 6. Heat map of the transitions between regulatory activities of the students in the "less successful" condition.
* TA: Task analysis; PL: Planning; EL: Elaboration; MO: monitoring; EX: Execution.

Table 5
Examples of sub-sequences frequently engaged in by the students in the "successful" and "less successful" conditions.

Performance Condition	Sub-sequence	Support Value	Count
Successful Condition	(EX) → (SSRL-MO)	0.82	81
	(SSRL-MO) → (EX)	0.81	80
	(EX) →(EX) → (SSRL-MO)	0.8	79
	(EX) → (SSRL-MO) → (SSRL-MO)	0.8	79
	(SSRL-MO) → (EX) → (EX)	0.8	79
	(SSRL-MO) → (SSRL-MO) → (SSRL-MO)	0.79	78
	(SSRL-MO) → (SSRL-MO) → (EX)	0.78	77
	(SSRL-MO) → (EX) → (EX) → (EX)	0.75	74
	(EX) → (EX) → (SSRL-MO) → (SSRL-MO)	0.74	73
	(EX) → (EX) → (EX) → (SSRL-MO)	0.73	72
	(SSRL-TA) → (SSRL-MO)	0.73	72
	(EX) → (SSRL-MO) → (EX)	0.72	71
Less Successful Condition	(SSRL-MO) → (EX)	0.71	29

* TA: Task analysis; MO: monitoring; EX: Execution.

task analysis. Of the self-regulated activities, the students were most likely to transition from SRL monitoring to SRL planning ($p = 0.44$). Of the socially shared regulatory activities, groups had a high probability of transitioning from SSRL monitoring to SSRL task analysis ($p = 0.36$), SSRL planning ($p = 0.30$), and SSRL elaboration ($p = 0.43$). The results also show a high probability of transitions from SSRL task analysis to SSRL planning ($p = 0.43$). Groups in the "less successful" condition also demonstrated a high probability of transitioning from SSRL monitoring to SRL task analysis ($p = 0.27$). These high probability transitions lay the foundation for revealing sequential patterns in regulatory activities which will be discussed in the next section.

Sequential pattern mining identified 355 subsequences for the "successful" condition and 327 subsequences with different support values for the "less successful" condition. Sub-sequences that contained only one regulatory activity were excluded from the final sequential patterns. Additionally, we only included sequential patterns that met a minimum support value of 0.7. As displayed in Table 5, the groups in the "successful" condition produced more significant patterns than did the groups in the "less successful" condition. In the "less successful" condition, (SSRL-MO) → (EX) was the only significant sub-sequences, revealing students' high reliance of SSRL monitoring to perform personal executing in solving tasks. In contrast, the groups in the "successful" condition regulated themselves in diverse ways, either by engaging in one or several executions immediately after engaging in SSRL monitoring (e.g., (SSRL-MO) → (EX) → (EX) → (EX)) or by engaging in one or several instances of SSRL monitoring immediately after engaging in execution (e.g., (EX) → (EX) → (SSRL-MO) → (SSRL-MO)). In addition, the subsequence (SSRL-TA) → (SSRL-MO) suggests that the groups in the "successful" condition were more likely to engage in group-level monitoring immediately after engaging in group-level task analysis.

5. Discussion

This study built a machine-learning model that semi-automatically identified self- and socially shared regulatory activities from the online chats and log files produced by students engaged in a computer-supported collaborative-learning STEM context. Specifically, we examined differences in the frequencies and variances of different regulatory activities and differences in sequences among them for successful and less successful groups. The results of this study reveal that techniques of temporal and sequential analysis (e.g. entropy analysis and sequential mining) can be used to identify differences in regulatory activities with increasing levels of granularity. Our findings will advance educational practices and the development of educational computer programs designed to better support collaborative learning.

In answer to our first research question, we found that the groups in the "successful" condition engaged in more regulatory activities—especially SRL monitoring, SRL elaborating, and SSRL task analysis—than did the groups in "less successful" condition. This result is consistent with the assumption that students who have better performance are more engaged in SRL processes (Sabourin, Mott, & Lester, 2013). It is possible that individuals who frequently regulated themselves to ensure good progress and who frequently used a formula to calculate the specific value of the resistor are more likely to perform better. Similar results have been obtained when students were asked to solve problems on their own (Taub, Azevedo, Bradbury, Millar, & Lester, 2018). The results of this study also align with the SRL model of Winne and Hadwin (1998) that students engaged in different SRL phases to different degrees. Specifically, students who performed better were more inclined to engage in elaboration as a means of strategic enactment in phase 3 and in monitoring to make adaptations in phase 4. Successful groups also contributed more effort in SSRL task analysis. This can be explained by the learning context, where no face-to-face environment was provided, and more group-level task analysis was needed to achieve mutual understanding. This finding can provide evidence for the study of Kapur, Voiklis, and Kinzer (2008), who found that exchanges of information (e.g. SSRL task analysis) in early stages can benefit collaborative learning.

Additionally, task execution, SRL monitoring, SSRL monitoring were the most frequent regulatory activities engaged in by the groups in both the "successful" and "less successful" conditions. The high frequency of task execution can be explained by the mechanisms of the STEM learning environment, in which the students had to engage in trial and error to solve the problem. The students spent significant time executing by testing their own ideas or following instructions from group members to adjust the resistors. SRL monitoring and SSRL monitoring were the second-most frequent regulatory activities, revealing the critical role of monitoring in connecting different regulatory activities and smoothing both SRL and SSRL processes. This result also relates to our findings for the third research question, which will be discussed below.

The results of the entropy analysis in the second research question revealed that the groups in the "successful" and "less successful" conditions differed significantly in the sequences of regulatory activities that they employed. The groups in "successful" condition had significantly lower entropy values and were more consistent in the regulatory processes that they employed, even though they displayed more regulatory activities. In contrast, the groups in "less successful" students had higher entropy values and transitioned frequently from one regulatory activity to another, even though they displayed fewer regulatory activities. This result suggests that the "successful" groups were more consistent in their use of regulatory strategies to accomplish a task. Instead of randomly trying different strategies, they may use specific regulatory strategies or follow specific regulatory mechanisms. This interesting finding supports the conclusion of Rogat and Linnenbrink-Garcia (2011) that synergy among regulatory processes improves the quality of collaborative learning by improving the productivity of a group. Dynamic synergy—or low levels of variation among self-regulatory and socially shared regulatory activities—is indicative of the effectiveness with which individuals in a group control themselves and group processes.

In answer to our third research question, we found differences in how the groups in the "successful" and "less successful" conditions transitioned between regulatory patterns. The successful groups started with task execution and ended with SSRL monitoring. They also engaged in more sequences incorporating execution and SSRL monitoring. In contrast, the less successful groups were most likely to start with task execution and end with task execution. They engaged in one sub-sequence: (SSRL-MO) → (EX). In agreement with the findings for the first research question, these findings further confirm the crucial role of SSRL monitoring in collaborative learning. Similar findings have also been obtained for face-to-face collaborative learning (Malmberg et al., 2017). The successful groups may benefit from the SSRL monitoring process by coordinating their interactions and resolving conflicts. This is why they started with execution and ended with SSRL monitoring. They took advantage of SSRL monitoring, either engaging in group-level monitoring after execution or engaging in execution after group-level monitoring. In contrast, the less successful students may have gotten stuck in execution; they started with execution, and they ended with execution. Monitoring can be engaged in at any time to promote knowledge construction and progress towards completing a task (Molenaar & Chiu, 2014; Sonnenberg & Bannert, 2015). This also suggests successful collaborative learners may be more aware of the timing of monitoring. They engage in socially shared monitoring whenever it is necessary to ensure progress in personal and group learning.

6. Implications for education

Findings of this study have important implications for STEM teaching and learning in CSCL environments. Firstly, since successful groups demonstrated more SRL and SSRL activities, it would be beneficial to encourage students to put more efforts in regulatory processes. The learning environment in this research sets an example for the design of CSCL environments for better support regulatory processes. For example, to facilitate a range of regulatory learning processes, the students were provided both individual goals and group goals. They were also given the option to track their personal learning progress and the progress of their group, allowing them to monitor and adjust their thinking and actions. For this reason, we highly recommend that teachers and computer programs establish CSCL environments that promote regulatory processes that improve group interactions and learning. Second, dynamic visualization of students' regulatory processes—especially one that included entropy values, which represent variation among regulatory activities—could provide instructors dynamic clues to students' learning trajectories. As Winne (2015) proposes, along with regulated learning, techniques of sequence mining or process mining could be used as a tool to manifest the process of group activities along with regulated learning. If an instructor was informed that the regulatory activities of a group were demonstrating little synergy, they could intervene to promote high-quality regulatory exchange (Rogat & Linnenbrink-Garcia, 2011). Finally, considering that socially shared monitoring dominates the regulatory processes and plays a significant role in promoting other regulation processes, educators should adaptively scaffold group learners by giving prompts regarding monitoring strategies or emphasize the importance of monitoring in their instructional design.

7. Conclusion, limitations, and future directions

The findings of our study add new evidence regarding the sequential patterns of self- and socially shared regulatory activities, particularly those that occur when students solve STEM tasks in CSCL environments. This study contributes to the literature on collaborative learning by: (1) highlighting the importance studying self-regulatory and socially shared regulatory activities simultaneously and (2) demonstrating variations in and sequential patterns of regulatory activities using different sequential mining techniques (e.g. entropy analysis, transition states, and sub-sequences). This study also further confirms the critical role of monitoring in directing other regulatory processes, and it proposes additional implications for teaching and learning in STEM and CSCL environments. However, a limitation of this study is that it considered only behavioral data (data from the log files) and data from online chats. Additional studies could replicate and extend the results of this study by examining other STEM tasks. Further research could also investigate whether students' prior knowledge or their motivation influence their regulatory processes and how monitoring might be triggered in SRL and SSRL. In addition, future studies could use other multi-channel data—e.g. data from think-aloud or from self-report questionnaires—to provide additional evidence regarding the self-level and group-level monitoring processes that students employ. It is also meaningful to further improve the SRL and SSRL automatic detection performance with more diverse features and advanced algorithms so that reliable automatic feedback can be provided to the groups on the fly. To extend our contribution to sequential mining, we will classify sequential patterns and further examine how SSRL monitoring plays a crucial role in CSCL.

Conflicts of interest

The authors declare that they have no conflict of interests.

Acknowledgement

Thanks to the Concord Consortium, especially Dr. Paul Horwitz and Cynthia McIntyre for supporting this work. We would also like to thank the children and teachers involving in this study for making it possible.

Appendix A. Supplementary data

Supplementary data to this article can be found online at https://doi.org/10.1016/j.compedu.2019.03.005.

Appendix I

In this section, three examples are selected in order to demonstrate how various features are extracted from the texts and further classified into certain SRL or SSRL categories. These categories are detected directly from the textual communication. We used term frequency-inverse document frequency (tf_idf) as a simple example to show how these features can be related with different coding categories. In reality, much more features are extracted and then input into more advanced machine learning algorithms to identify the SRL and SSRL categories.

SRL-Task Analysis provides the understanding about the purposes of the tasks based on the previous self-knowledge
Eg. "Gee, each of those steps could have been 'broken down' or sub-divided into even smaller steps, I am thinking about that I need to do it again."
SSRL-Task analysis provides collaboration information about doing projects in groups
Eg. "These parts of the plane can not be separated, how can we put them together."
Execution provides the information of actions related to project
Eg. "The bugs related to the delivery orders module were removed"

In order to relate the word feature in the above examples to the SRL and SSRL categories, the following three steps are performed:

Step 1 After the training corpus was developed based on the coding framework, stopwords were removed and tokenized corresponding to the SRL and SSRL type shown in Table 1. So we have token corpus that can best represent the SRL and SSRL type;
Step 2 Calculate the frequency of the words in the corpus that appeared in the target text;
Step 3 The similarity between the target text tokens and corpus tokens were calculated based on which we can classify the target text into a specific SRL and SSRL class.
The following is the example of classifying a text into a specific SRL and SSRL type based on the tf-idf (term frequency-inversed document frequency) technique.
To illustrate, two document copus belongs to SRL_Task Analysis and SRL_Planning respectively are shown in the example, and each copus including 1000 documents. We want to extract the SRL and SSRL contains in the text of "*Gee, each of those steps could have been 'broken down' or sub-divided into even smaller steps, I am thinking about that I need to do it again.*"
According to step1, we have the corpus tokens, in order to make the demonstration easier, we do the following assumptions:
As for the SRL_Task Analysis, we set the extracted tokens of the corpus based on the introduced method as:
Lingustic Feature: *sentence length (in this paper we count the number of different words in a sentence as the length of sentence);*
LIWC-related features: "*I*", "*could*", "*Gee*", "*need*",*do*","*think*","*try*","*understand*";
Latent Dirichlet allocation (LDA) features: "*task*","*content*";
Assume each of these 10 words and each of them appear in 10 of 1000 documents.
Similarly for the SRL_Planning, we set the extracted tokens of the corpus as
Lingustic Feature: *sentence length (in this paper we count the number of different words in a sentence as the length of sentence);*
LIWC-related features: "*let*","*me*","*divided*","*hold*","*need*","*solve*","*I*","*think*";
LDA features: "*task*","*readjust*";
Where, each of these 10 words and each of them appear in 10 of 1000 documents.
Then, calculate the term frequency-inverse document frequency(tf_idf) for each words in terms of the two classes.
tf_idf(*I*) = 2/25*log(1000/10)
tf_idf(*could*) = 1/25*log(1000/10)
tf_idf(*need*) = 1/25*log(1000/10)
tf_idf(*Gee*) = 1/25*log(1000/10)
tf_idf(*do*) = 1/25*log(1000/10)

tf_idf(*task*) = 0
tf_idf(*think*) = 1/25*log(1000/10)
tf_idf(*try*) = 0
tf_idf(*understand*) = 0
tf_idf(*base*) = 0
tf_idf(*content*) = 0
tf_idf(*divided*) = 1/25*log(1000/10)
Finally, sum the tf_idf of the words that appears in the target text respectively, and assign the SRL and SSRL class to it according the larger sum of tf_idf value.
Sum(tf_idf_Task) = tf_idf(I) + tf_idf(could) + tf_idf(need) + tf_idf(Gee) + tf_idf(do) + tf_idf(think) = 7/25*2 = 0.56
Sum(tf_idf_Planning) = tf_idf(I) + tf_idf(think) + tf_idf(need) + tf_idf(divided) = 0.4
Sum(tf_idf_Task) > Sum(tf_idf_Planning), so we can assign the target task to the SRL Task Analysis category.

References

Azevedo, R. (2014). Issues in dealing with sequential and temporal characteristics of self- and socially-regulated learning. *Metacognition and Learning, 9*(2), 217–228.
Bannert, M., Reimann, P., & Sonnenberg, C. (2014). Process mining techniques for analysing patterns and strategies in students' self-regulated learning. *Metacognition and Learning, 9*(2), 161–185.
Bruno, R. J. (2010). *Social differentiation, participation inequality and optimal collaborative learning online.* Unpublished doctoral thesisWest Lafayette, USA: Purdue university.
Dalal, M. K., & Zaveri, M. A. (2011). Automatic text classification: A technical review. *International Journal of Computer Application, 28*(2), 37–40.
Hadwin, A. F., Järvelä, S., & Miller, M. (2011). Self-regulated, co-regulated, and socially shared regulation of learning. *Handbook of self-regulation of learning and performance, 30*, 65–84.
Isohätälä, J., Järvenoja, H., & Järvelä, S. (2017). Socially shared regulation of learning and participation in social interaction in collaborative learning. *International Journal of Educational Research, 81*, 11–24. https://doi.org/10.1016/j.ijer.2016.10.006.
Järvelä, S., & Hadwin, A. F. (2013). New frontiers: Regulating learning in CSCL. *Educational Psychologist, 48*(1), 25–39.
Jarvela, S., Kirschner, P. A., Hadwin, A., Jarvenoja, H., Malmberg, J., Miller, M., et al. (2016). Socially shared regulation of learning in CSCL: Understanding and prompting individual- and group-level shared regulatory activities. *International Journal of Computer-Supported Collaborative Learning, 11*(3), 263–280.
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, 42*, 1–11. https://doi.org/10.1016/j.learninstruc.2015.10.006.
Jarvela, S., Volet, S., & Jarvenoja, H. (2010). Research on motivation in collaborative learning: Moving beyond the cognitive-situative divide and combining individual and social processes. *Educational Psychologist, 45*(1), 15–27.
Kapur, M., Voiklis, J., & Kinzer, C. K. (2008). Sensitivities to early exchange in synchronous computer-supported collaborative learning (CSCL) groups. *Computers & Education, 51*(1), 54–66.
Kotsiantis, S. B., Zaharakis, I., & Pintelas, P. (2007). Supervised machine learning: A review of classification techniques. *Emerging artificial intelligence applications in computer engineering, 160*, 3–24.
Kwon, K., Liu, Y. H., & Johnson, L. P. (2014). Group regulation and social-emotional interactions observed in computer supported collaborative learning: Comparison between good vs. poor collaborators. *Computers & Education, 78*, 185–200. https://doi.org/10.1016/j.compedu.2014.06.004.
Lajoie, S. P., & Lu, J. (2012). Supporting collaboration with technology: does shared cognition lead to co-regulation in medicine? *Metacognition and Learning, 7*(1), 45–62.
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*, 160–174. https://doi.org/10.1016/j.cedpsych.2017.01.009.
Matei, S., Oh, K., & Bruno, R. (2006). Collaboration and communication: A social entropy approach. *Proceedings of National Communication Association Annual Convention, San Antonio*.
McCardle, L., & Hadwin, A. F. (2015). Using multiple, contextualized data sources to measure learners' perceptions of their self-regulated learning. *Metacognition and Learning, 10*(1), 43–75.
Miller, M., Järvelä, S., & Hadwin, A. (2017). *Self-regulation, co-regulation, and shared regulation in collaborative learning environments Handbook of self-regulation of learning and performance.* Routledge99–122.
Molenaar, I., & Chiu, M. M. (2014). Dissecting sequences of regulation and cognition: Statistical discourse analysis of primary school children's collaborative learning. *Metacognition and Learning, 9*(2), 137–160.
Mu, J., Stegmann, K., Mayfield, E., Rose, C., & Fischer, F. (2012). The ACODEA framework: Developing segmentation and classification schemes for fully automatic analysis of online discussions. *International Journal of Computer-Supported Collaborative Learning, 7*(2), 285–305.
Panadero, E., & Jarvela, S. (2015). Socially shared regulation of learning: A review. *European Psychologist, 20*(3), 190–203.
Pennebaker, J. W., Boyd, R. L., Jordan, K., & Blackburn, K. (2015). *The development and psychometric properties of LIWC2015.* (Retrieved from).
Pintrich, P. R. (2000). *The role of goal orientation in self-regulated learning Handbook of Self-Regulation.* Elsevier451–502.
Popov, V., Xing, W., Zhu, G., Horwitz, P., & McIntyre, C. (2018, June). The influence of students' transformative and non-transformative contributions on their problem solving in collaborative inquiry learning. In J. Key, & R. Luckin (Vol. Eds.), *Rethinking Learning in the Digital Age. Making the Learning Sciences Count: Procceedings of The International Conference of the Learning Sciences (ICLS, 2018): Vol. 3*, (pp. 855–862). London, UK: International Society of the Learning Sciences.
Reimann, P., Markauskaite, L., & Bannert, M. (2014). e-Research and learning theory: What do sequence and process mining methods contribute? *British Journal of Educational Technology, 45*(3), 528–540.
Rogat, T. K., & Linnenbrink-Garcia, L. (2011). Socially shared regulation in collaborative groups: An analysis of the interplay between quality of social regulation and group processes. *Cognition and Instruction, 29*(4), 375–415.
2013//Sabourin, J., Mott, B., & Lester, J. (2013). Discovering behavior patterns of self-regulated learners in an inquiry-based learning environment. *Paper presented at the artificial intelligence in education, Berlin, Heidelberg*.
Soller, A. (2004). Computational modeling and analysis of knowledge sharing in collaborative distance learning. *User Modeling and User-Adapted Interaction, 14*(4), 351–381. https://doi.org/10.1023/B:USER.0000043436.49168.3b.
Sonnenberg, C., & Bannert, M. (2015). Discovering the effects of metacognitive prompts on the sequential structure of SRL-processes using process mining techniques. *Journal of Learning Analytics, 2*(1), 72–100.
Splichal, J. M., Oshima, J., & Oshima, R. (2018). Regulation of collaboration in project-based learning mediated by CSCL scripting reflection. *Computers & Education, 125*, 132–145.
Taub, M., Azevedo, R., Bradbury, A. E., Millar, G. C., & Lester, J. (2018). Using sequence mining to reveal the efficiency in scientific reasoning during STEM learning with a game-based learning environment. *Learning and Instruction, 54*, 93–103.
Vauras, M., Iiskala, T., Kajamies, A., Kinnunen, R., & Lehtinen, E. (2003). Shared-regulation and motivation of collaborating peers: A case analysis. *Psychologia, 46*(1), 19–37.
Volet, S., Summers, M., & Thurman, J. (2009). High-level co-regulation in collaborative learning: How does it emerge and how is it sustained? *Learning and Instruction,*

19(2), 128–143.
Volet, S., Vauras, M., & Salonen, P. (2009). Self- and social regulation in learning contexts: An integrative perspective. *Educational Psychologist, 44*(4), 215–226.
FebruaryWang, Y.-C., Kraut, R., & Levine, J. M. (2012). To stay or leave?: The relationship of emotional and informational support to commitment in online health support groups. *Paper presented at the 2Proceedings of the ACM 2012 conference on computer supported cooperative work*.
Winne, P. H. (2015). What is the state of the art in self-, co- and socially shared regulation in CSCL? *Computers in Human Behavior, 52*, 628–631. https://doi.org/10.1016/j.chb.2015.05.007.
Winne, P. H., & Hadwin, A. F. (1998). Studying as self-regulated learning. *Metacognition in educational theory and practice, 93*, 27–30.
Winne, P. H., Hadwin, A. F., & Perry, N. E. (2013). Metacognition and computer-supported collaborative learning. *International handbook of collaborative learning,* 462–479.
Winne, P. H., & Jamieson-Noel, D. (2002). Exploring students' calibration of self reports about study tactics and achievement. *Contemporary Educational Psychology, 27*(4), 551–572.
Winne, P. H., & Perry, N. E. (2000). In M. Boekaerts, & P. R. Pintrich (Eds.). *Measuring self-regulated learning* (pp. 531–566). Academic Press.
Xing, W. (2017). Exploring the knowledge curation process of online health communities. *The IEEE Intelligent Informatics Bulletin, 18*(1), 27–28.
Xing, W., & Gao, F. (2018). Exploring the relationship between online discourse and commitment in Twitter professional learning communities. *Computers & Education, 126*, 388–398.
Xing, W., Goggins, S., & Introne, J. (2018). Quantifying the effect of informational Support on membership retention in online communities through large-scale data analytics. *Computers in Human Behavior, 86*, 227–234.
Zimmerman, B. J. (2000). Attaining self-regulation: A social cognitive perspective. *Handbook of Self-Regulation,* 13–39. https://doi.org/10.1016/B978-012109890-2/50031-7.
Zimmerman, B. J. (2008). Investigating self-regulation and motivation: Historical background, methodological developments, and future prospects. *American Educational Research Journal, 45*(1), 166–183. https://doi.org/10.3102/0002831207312909.
Zhu, G., Xing, W., & Popov, V. (2019). Uncovering the sequential patterns in transformative and non-transformative discourse during collaborative inquiry learning. *The Internet and Higher Education, 41*, 51–61.

Juan Zheng is a PhD candidate and research assistant at Educational Counselling and Psychology (ECP), McGill University with background in educational technology and learning sciences. Her research interests are self-regulated learning, academic emotions, and educational data mining

Wanli Xing is an Assistant Professor in Instructional Technology at Texas Tech University, USA with background in learning sciences, statistics, computer science and mathematical modeling. His research interests are educational data mining, learning analytics, and CSCL.

Gaoxia Zhu is a PhD candidate and research assistant in the Institute for Knowledge Innovation & Technology, Ontario Institute for Studies in Education (OISE), University of Toronto with background in Educational Technology and Curriculum Studies. Her research interests include Knowledge Building, learning analytics, and CSCL.