JOURNAL OF LEARNING ANALYTICS

Volume 8(1), 49–74. http://doi.org/10.18608/jla.2021.7230

Examining Student Regulation of Collaborative, Computational, Problem-Solving Processes in Open-Ended Learning Environments

Mona Emara[1], Nicole M. Hutchins[2], Shuchi Grover[3], Caitlin Snyder[4], Gautam Biswas[5]

Abstract

The integration of computational modelling in science classrooms provides a unique opportunity to promote key 21st century skills including computational thinking (CT) and collaboration. The open-ended, problem-solving nature of the task requires groups to grapple with the combination of two domains (science and computing) as they collaboratively construct computational models. While this approach has produced significant learning gains for students in both science and CT in K–12 settings, the collaborative learning processes students use, including learner regulation, are not well understood. In this paper, we present a systematic analysis framework that combines natural language processing (NLP) of collaborative dialogue, log file analyses of students' model-building actions, and final model scores. This analysis is used to better understand students' regulation of collaborative problem solving (CPS) processes over a series of computational modelling tasks of varying complexity. The results suggest that the computational modelling challenges afford opportunities for students to a) explore resource-intensive processes, such as trial and error, to more systematic processes, such as debugging model errors by leveraging data tools, and b) learn from each other using socially shared regulation (SSR) and productive collaboration. The use of such SSR processes correlated positively with their model-building scores. Our paper aims to advance our understanding of collaborative, computational modelling in K–12 science to better inform classroom applications.

Notes for Practice

- This paper provides a framework that combines log data analyses and natural language processing to understand how CPS and regulation are employed in tasks of varying levels of difficulty and scaffolding.
- Results indicate that students engaged in more socially shared regulation (SSR) and productive collaboration in more challenging, open-ended tasks than in scaffolded tasks. SSR also correlated with more productive co-construction of knowledge, leading to higher performance scores.
- Our findings help build a better understanding of regulation processes and co-construction of knowledge during computational modelling in K–12 science. This understanding has the potential to inform the design of future environments and tasks to foster better collaboration and learning in computational scientific modelling and beyond.

Keywords

Collaborative problem solving, socially shared regulation, natural language processing, process mining, computational modelling

Submitted: 09/06/20 — **Accepted:** 23/01/21 — **Published:** 09/04/21

Corresponding author [1]Email: mona.emara@dmu.edu.eg Address: Department of Educational Psychology, Faculty of Education, Damanhour University, Alabadya, Damanhour 22516, Egypt. ORCID ID: https://orcid.org/0000-0002-9277-7073

[2]Email: nicole.m.hutchins@vanderbilt.edu Address: Institute for Software Integrated Systems, School of Engineering, Vanderbilt University,1025 16th Ave S, Nashville, TN 37212, USA. ORCID ID: https://orcid.org/0000-0002-7258-5023

[3]Email: shuchig@cs.stanford.edu, Address: Looking Glass Ventures, Palo Alto, CA, USA. ORCID ID: https://orcid.org/0000-0001-6633-8862

[4]Email: caitlin.r.snyder@vanderbilt.edu, Address: Institute for Software Integrated Systems, School of Engineering, Vanderbilt University, 1025 16th Ave S, Nashville, TN 37212, USA.

[5]Email: gautam.biswas@vanderbilt.edu, Address: Institute for Software Integrated Systems, School of Engineering, Vanderbilt University, 1025 16th Ave S, Nashville, TN 37212, USA. ORCID ID: https://orcid.org/0000-0002-2752-3878

1. Introduction

Driven by the needs of 21st century education, stakeholders recognize that collaborative problem solving (CPS) and computational thinking (CT) are increasingly important skills for academic and career success for all students (Grover & Pea, 2013). CT skills, which collectively encapsulate a set of problem-solving strategies, represent an emerging area of interest in all K–12 science, technology, engineering, and mathematics (STEM) disciplinary work today (Henderson, Cortina, & Wing, 2007). Interest in integrating CT in STEM learning is further motivated by current STEM workforce practices that increasingly rely on computational modelling and simulation tools for understanding, analyzing, designing, and solving problems (Grover & Pea, 2013; Landau, 2006). However, students face a number of challenges during computational modelling that include applications of CT concepts (e.g., conditional logic, loops) and practices (e.g., translating domain knowledge into computational form, debugging). There has been insufficient attention paid to understanding how students regulate their learning to overcome these difficulties.

Evaluations of CPS during computational modelling have been an informative source for increasing our understanding of the regulation processes students implement to complete complex model-building tasks (Emara, Grover, Hutchins, Biswas, & Snyder, 2020; Grover, Hutchins, Biswas, Snyder, & Emara, 2019). Preliminary efforts targeting deeper understanding of CPS during computational modelling have demonstrated the impact of regulation of CPS in helping students overcome known computational difficulties, including debugging and applications of conditional logic (Emara et al., 2020). In addition, our prior analysis of collaborative discourse utilizing a CPS framework has increased our understanding of how students combine domain and CT concepts and practices to construct models and solve problems (Snyder et al., 2019). However, the measurement of CPS during computational modelling thus far has often focused on individual learning gains rather than examining the actual collaborative learning processes (e.g., Loksa & Ko, 2016; Peters-Burton, Cleary, & Kitsantas, 2018). Existing research recognizes the critical role played by learning analytics approaches to interpret student learning behaviours and learning processes during collaboration (Kapur, 2016; Soderstrom & Bjork, 2015). This includes the primary motivation for this paper, i.e., analyzing the content and linguistic features of student discourse linked to their knowledge construction and interaction processes using natural language processing (NLP) approaches (Fischer et al., 2020). Our work aims to contribute a systematic approach for understanding the regulation of CPS processes as students co-construct STEM computational models in an open-ended learning environment (OELE). Extending our previous work in which we implemented a coding scheme for the evaluation of verbal data (Emara et al., 2020), this work leverages the ideas of linguistic modelling (Ferreira, Kovanović, Gašević, & Rolim, 2018) to integrate our coding scheme with NLP methods and computational modelling process analysis of trace action data during model-building to target the following research questions:

RQ1. What is the nature of CPS regulation activated by students when they work in groups on three different types of physics modelling tasks of varying complexity?

RQ2. How do students' self- and shared regulatory activities correlate with their performance (model-building scores)?

RQ3. How do action patterns and problem-solving strategies derived using NLP emerge across collaborative computational modelling tasks of varying difficulty?

There is increasing interest in designing adaptive support and timely feedback for CPS (Noroozi et al., 2019; Sobocinski et al., 2020). However, for adaptive collaborative modelling in OELEs, this support requires a better understanding of the regulatory processes underlying collaboration to provide the support when it is needed. The understanding gained through our analyses and answering of these research questions will aid the design of future learning interventions and collaborative OELEs.

The remaining sections of this paper are organized as follows. In Section 2 we provide background literature on computational modelling and the regulation of CPS as well as measurement methods and analyses to provide a framing for our analysis approach and demonstrate how we extend the literature. In Section 3 we discuss the study and the Collaborative, Computational STEM (C2STEM) OELE, followed by the detailing of our analysis approach in Section 4. Section 5 presents our results and Sections 6 and 7 conclude the paper with a discussion on our findings, limitations, and future work.

2. Background and Related Work

2.1. Computational Modelling in Support of Synergistic Physics and Computational Thinking (CT) Learning

Our work focuses on the evolution of students' regulation of problem-solving processes across a sequence of computational modelling tasks. In our environment, C2STEM, students construct computational models using block-based, domain-specific modelling languages (DSMLs) to represent the behaviour of relevant scientific phenomena (Hutchins et al., 2020a). The environment also provides a discrete-time (step-by-step) simulation method that students can use to execute their models and analyze the behaviours generated by their models (Hutchins et al., 2020b). The learning-by-modelling framework illustrated in Figure 1, adapted from Hutchins et al. (2020a), demonstrates the subprocesses that students may employ to build computational models. Research has demonstrated the effectiveness of STEM as a vehicle to support the learning and understanding of computational concepts and practices (Papert & Harel, 1991; Hutchins et al., 2019; Sengupta, et al., 2013; Weintrop et al., 2016). In addition, computational models and CT have proven to be effective tools for the learning of difficult science concepts and practices (Hambrusch et al., 2009; Jona et al., 2014; Repenning et al., 2010), especially when introduced through discrete and qualitative forms of fundamental laws, instead of equation-based continuous forms (Redish & Wilson, 1993). This simultaneous benefit, or synergistic learning, is predicated on the idea that the dual engagement in STEM and CT extends opportunities to simultaneously apply and learn constructs of the target domains.

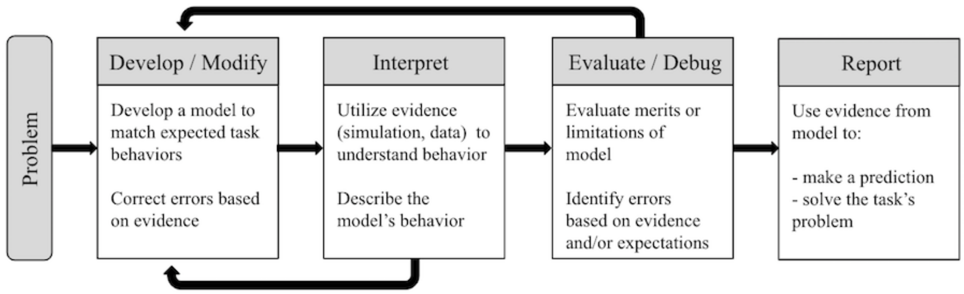


Figure 1. Processes and subprocesses integral for learning-by-modelling (Hutchins et al., 2020a).

Previous research has revealed some of the difficulties that students face in computational modelling tasks when they work individually, including difficulties translating STEM concepts and practices into computational form (Basu, Biswas, & Kinnebrew, 2016), debugging errors in developed computational models (Hutchins et al., 2020b), and explaining model behaviour based on model evidence (Grover et al., 2019). To evolve our understanding of these difficulties, incorporating research that targets CPS and examines student collaborative regulation of problem-solving skills may allow us to examine real-time processes that students employ to solve complex computational problems. For instance, Grover et al. (2019) demonstrated how a coordinated examination of regulation behaviours with an examination of student applications of STEM and CT can provide useful insight into how students debug their computational models, initialize appropriate STEM variables, and implement conditional logic. In the following sections, we will provide background supporting our CPS and regulation analysis approach in the context of computational modelling in STEM.

2.2. Understanding the Regulation of CPS Processes During Computational Modelling

To understand the regulation of CPS processes during co-construction of STEM computational models in an OELE, we derive our theoretical framework from key self- and socially shared regulation frameworks (Hadwin et al., 2018; Winne & Hadwin, 1998). Models of regulation explain learning through different phases — often called micro-level processes. When solving problems, self-regulated learners use cognitive processes (e.g., read, represent, test) to solve their problem and metacognitive processes (plan, monitor, and evaluate) to control and monitor their problem solving. These learners often learn more than other students who do not engage in these regulation processes (Greene & Azevedo, 2009; Hadwin et al., 2018; Klahr & Carver, 1988; Winne & Hadwin, 1998).

In collaborative problem solving, regulation happens both on an individual (SR) and on a group level (socially shared regulation; Hadwin et al., 2018; Wise et al., 2015). Socially shared regulation (SSR) can establish clear expectations and encourage group members to reflect on and combine their strategies and abilities in productive ways (Xie et al., 2018). Additionally, SSR can provide examples of productive collaboration through effective strategies like questioning, encouraging contributions from others, and clarifying or elaborating ideas (De Backer et al., 2015). However, students often experience difficulties in adequately regulating their problem solving during computational modelling by jointly adjusting their metacognitive and behavioural states as they translate their STEM knowledge into computational model representations (e.g., Hutchins et al., 2020a, 2020b; Sengupta, et al., 2013). Recently, there has been a lot of interest in understanding of regulation processes during CPS, especially during collaborative scenarios of different complexity (Noroozi et al., 2019; Sobocinski et al., 2020; Sun, Shut, Stewart, Yonehir, Duran, & D'Mello, 2020). However, there is a lack of clarity on how students' productive and unproductive discussions interplay with micro-level cognition and metacognition regulation processes during collaborative, computational modelling in OELE. We believe that examining collaborative open-ended, problem solving during computational modelling through the lens of SR and SSR will provide insights into student collaboration as well as the learning of science and CT.

2.3. Analyzing Regulation of CPS Processes During Computational Modelling

Collaboration during science simulations has been studied through collaborative discourse (Roschelle & Teasley, 1995; Gobert et al., 2007). Analyzing regulation activities in the discourse data bears the potential to not only be a more reliable measurement but also a better predictor than self-reports or interviews (Molenaar & Chiu, 2014; Schoor & Bannert, 2012).

Researchers are beginning to explore uses of educational data mining and learning analytics to understand regulation of CPS processes (e.g., Dowell, Lin, Godfrey, & Brooks, 2020; Yett et al., 2020). Machine learning techniques can be used to deepen insights about how students engage in regulation of CPS processes by analyzing sequences of learning activities (Siadaty et al., 2016; Gašević et al., 2017). Most studies of regulation of CPS that applied machine learning used raw trace data derived from log-files as the main source of evidence (e.g., Järvelä, Malmberg, & Koivuniemi, 2016; Schoor & Bannert, 2012). Recently, studies have attempted to assess student self-regulation processes while debugging by analyzing individuals' think alouds combined with trace methodologies (e.g., Lin et al., 2015; Loksa & Ko, 2016) or a multiple-choice test where learners interacted with a computer agent (Liu et al., 2017). Limited research has combined collaborative discourse and trace action data in the context of open-ended problem solving to investigate differences in regulatory processes and learning behaviours.

Finally, we aim to investigate student discourse and actions through the dual lenses of cognitive and metacognitive processes to understand the evolution of regulation processes of CPS during computational modelling in science. The benefits of evaluating the regulation of CPS processes over an extended period (across tasks) may offer new insights into understanding changes in computational modelling processes as well as the impact of computational modelling task difficulty on regulation. For example, Malmberg et al. (2014) provide empirical evidence showing students using different types of self-regulatory activities during different challenging problems.

2.4. Natural Language Processing (NLP) and CPS

Our work is grounded in the research on linguistic modelling of CPS processes and outcomes, such as knowledge co-construction (Weinberger & Fischer, 2006), argumentation (Rosé, et al., 2008), and task performance (Amon, Vrzakova, & D'Mello, 2019). A large majority of NLP research has focused on surface-level text processing (e.g., pitch, tone, and/or turn-taking) to help interpret group functioning, and the available tools consequently emphasize the central role of accurate word- and sentence-level text processing (Praharaj, Scheffel, Drachsler, & Specht, 2018) rather than the semantic meaning of the utterances. Within the context of learning analytics of computational modelling, we aim to integrate NLP to support the semantic analysis of human dialogue. Thus, we focus on tools developed to calculate linguistic indices that move beyond these surface-level tasks and provide information that may be more important within educational contexts. Multiple characteristics of language can be gleaned from the words (including n-grams and POS) and captured using both techniques for analyzing observable features (e.g., word frequencies, word-document distributions) and latent meaning from the text (McNamara, 2011).

Recently, there has been a growing effort to apply NLP analytics strategies to study learner roles and regulation in collaborative discourse in terms of word, sentence, and paragraph counts; word cloud visualization; sentiment analysis; and lexical diversity type-token ratio calculations to determine text cohesion. For example, Dowell et al. (2020) applied a computational linguistic framework to analyze the sequential interactions of online team communication and to detect roles in regulation, social coordination, and meaning-making in discussion. In addition, Sullivan and Keith (2019) used a parts-of-speech (POS) tagging program to automatically parse a transcript of spoken dialogue collected from a small group of middle school students involved in solving a robotics challenge. They grammatically analyzed the dialogue at the level of the tri-gram. They then interpreted the POS tri-grams within the theoretically derived actions and objects in their specific robotics problem space.

Our research focuses on analyzing discourse data about computational modelling captured on video and audio in our learning-by-modelling environment, C2STEM. Our work is comparable to that of other educational researchers who have focused frequencies of n-grams (i.e., word sequences of length n) in verbal language such as Sullivan and Keith (2019), Worsley and Blikstein (2011), and Stewart et al. (2019). However, our study differs from these efforts in that we seek to understand unprompted regulation of CPS conversations in-situ and in-process during computational modelling in STEM learning. More interestingly, we suggest a combined use of two well-established analytic techniques in the field of learning analytics — process mining (PM) and NLP — to define how groups adapt their behaviours and explore different options in their collaboration when confronted with challenges. Although both techniques have been used for analysis of collaborative learning, their combined use has been limited. This paper demonstrates how the two methods can complement each other in the analysis of collaborative learning in the context of an OELE. In the following section, we detail our collaboration study, the C2STEM OELE, and our data collection process.

3. The Study

3.1. Participants and Setting

The study was conducted in a high school classroom run on a university campus in the Southern United States. The students were participants in a selective program designed to immerse high school students in advanced academic experiences in a university setting. These students previously completed a 4-week, daily summer immersion program in which they participated in collaborative projects and activities; therefore, they had prior experience working with each other. Fourteen 9th grade students (six female and eight male; mean age 15 years) were assigned to five groups. The grouping yielded one all-male dyad and four mixed-gender triads. The groups were formed by a member of the research team who had worked with some of the participants previously. All groups contained a student with prior C2STEM experience and the remaining students were randomly assigned.

3.2. C2STEM Environment and Tasks

The C2STEM environment uses NetsBlox (Broll et al., 2017), an extension of Snap!,[1] with custom domain-specific blocks (e.g., blocks for setting and updating *position*, *velocity*, *acceleration*, and *heading*) that help learners focus on physics concepts. Groups worked on our curriculum once a week for two months. Following a 45-minute training unit that introduced students to C2STEM, students worked on two C2STEM modules that covered 1D motion with acceleration and 2D motion with constant velocity. Our curriculum is scaffolded by types of tasks to support a progression in complexity of STEM and CT concepts and practices and is described in detail in Hutchins et al. (2020b). In our analyses, we focus on instructional and challenge tasks. Instructional tasks centre on domain (physics) concepts, with minimal CT applications (especially applications of known difficulties such as conditional logic), and represent the lowest complexity for our analyses. Although minimal instruction is provided (for instance, students are tasked to "Simulate the motion of the sloth moving to the right with a constant velocity of 10 m/s starting at a position of 0 m"), instructional tasks are the first implementation of newly introduced STEM constructs and we hypothesize that previously identified difficulties such as the translation of STEM into computational form (Basu et al., 2016) may be exacerbated at this time. The *challenge tasks* require more advanced CT applications with students having to critically think about the program structure and CT to successfully solve the problem. For the example group solution for the 2D motion challenge task in Figure 2, students were tasked to "Simulate the motion of the boat crossing the river, stopping at both islands along the way." Challenge task complexity increased slightly from the 1D challenge task to the 2D challenge task. As the 2D challenge task was implemented at the end of the second physics unit, we hypothesize that with adequate knowledge gains in physics and CT over the course of the unit (with an average of three tasks prior to the challenge), collaborative problem-solving approaches will improve (e.g., abilities to develop a shared understanding of expectations and implementations of the model).

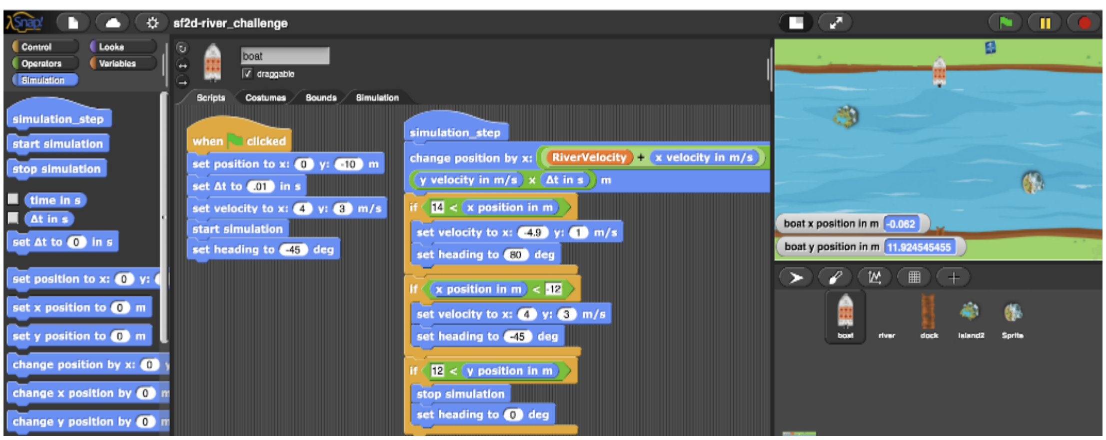


Figure 2. Example 2D constant velocity student group solution.

3.3. Data Collection

Each group worked on a single computer. Students in the group were instructed to switch between "driver" and "navigator" roles (Williams et al., 2002) between tasks. Visual and verbal behaviour was recorded using OBS™ screen-capture software that recorded mouse movements, video, and audio. All sessions for the five groups performing the three tasks were videotaped, resulting in fifteen sessions with 11 hours of video recordings. In addition, one camera was used to capture pictures and sound for all students. The recorded sessions provided insight into group regulation behaviours across the three problems in the C2STEM environment.

4. A Multimodal Learning Analytics (MMLA) Approach to Evaluating CPS

Our efforts target a deeper analysis of CPS through the mapping of logged model-building action sequences with group discourse using an NLP approach to collaborative discourse analysis. The goal is to evaluate how approaches to CPS as

[1] http://snap.berkeley.edu/

evidenced through group discussions may impact model-building processes and vice versa. Figure 3 illustrates our analysis framework.

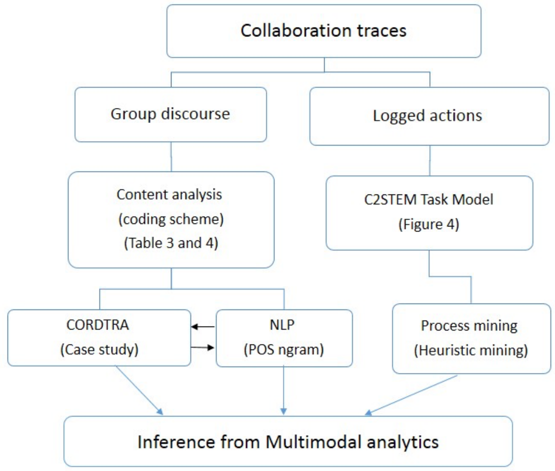


Figure 3. The methodological steps followed for the combined use of PM and NLP.

In order to accomplish this, we 1) adopt a theory- and data-driven framework for analyzing and interpreting student actions based on a hierarchical task model (cf. Grover et al., 2017; Kinnebrew, Segedy, & Biswas, 2017); 2) apply our coding scheme for student regulation processes (Emara, Tscholl, Dong, & Biswas, 2017; Emara et al., 2020); 3) implement an exploratory NLP approach to identify and evaluate CPS topics (e.g., explanatory vs. instructional discourse) based on our prior human coding and analysis work; and 4) evaluate student learning using an evidence-based rubric that captures the STEM and CT concepts and practices required for each computational modelling task. Table 1 provides hypothesized applications of productive and unproductive collaboration during computational modelling that will be tested and evaluated in our analyses. In the following subsections, we will detail our approach for each type of analysis, and refer back to previous work that labels these processes as "productive" or "unproductive."

Table 1. Framing Our Multimodal Collaborative, Computational Modelling Analysis Approach

Collaboration	STEM+CT	Regulation	Model-Building Processes	CPS Discourse
Productive	High task scores	Predominant socially shared regulation Evidence of model planning, reflection	Processes indicate systematic implementation of actions (e.g., running simulation and evaluating data tools to support error identification)	Use of explanatory words such as "because," "think" Use of conversation guiding words such as "how"
Unproductive	Difficulties in STEM and CT, or both as evidenced by task scores	Predominant self-regulation Lack of problem explanation or interpretation	Processes do not demonstrate systematicity and/or represent a more trial and error approach	Use of instructional verbiage such as "do [action]"

4.1. Process Mining

Interpreting actions is more productive if we can associate them with the specific goals (i.e., the context) that students may have when performing a set of actions. For our analysis, the student actions were represented at a level of abstraction, so that action patterns could be derived and semantically interpreted in terms of students' model-building processes (Segedy, Kinnebrew, & Biswas, 2015; Werner et al., 2013). In order to represent student actions in C2STEM at a level of abstraction that makes it easier to interpret their modelling action sequences, following previous work (e.g., Hutchins, Biswas, Grover, Basu, & Snyder, 2019), we created the task model illustrated in Figure 4. We extracted and interpreted student action sequences that are linked to code construction and code evaluation and can be linked to key subprocesses described in Figure 1 (e.g., "Identify model errors based on evidence and expectations"). For example, students may execute or "PLAY" their model code to see (evaluate) if the generated motion matches their intuition of the correct motion of the object. We provide example screenshots of actions from the Task model (Table 2) for context supporting interpretation of the results.

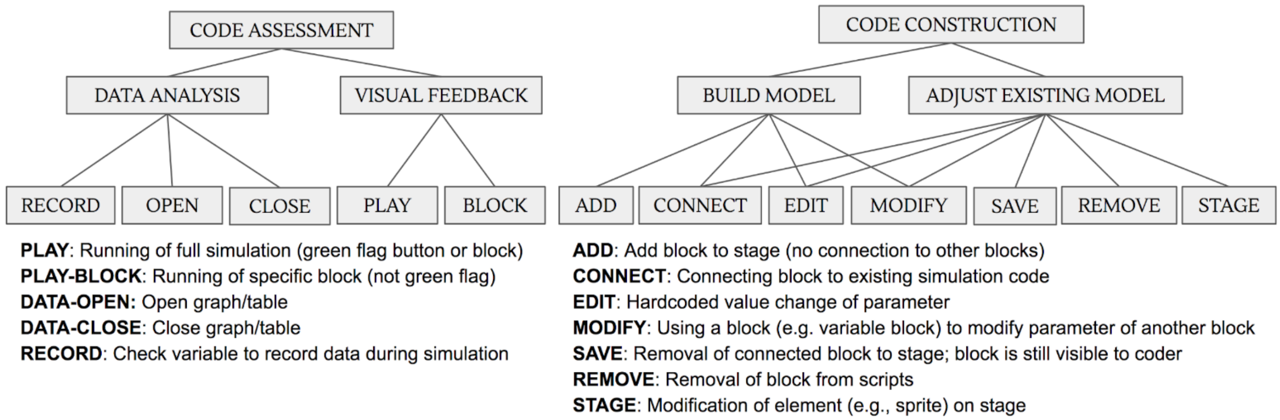
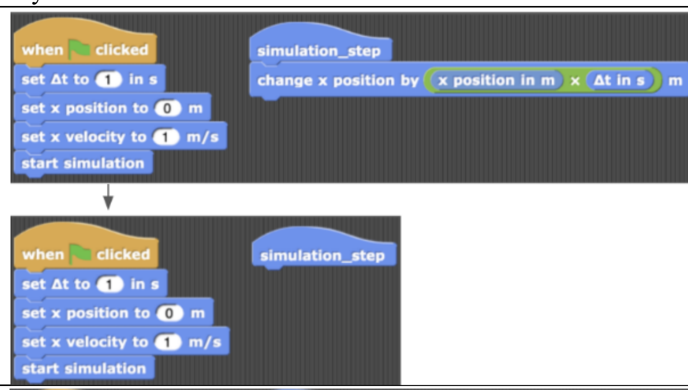
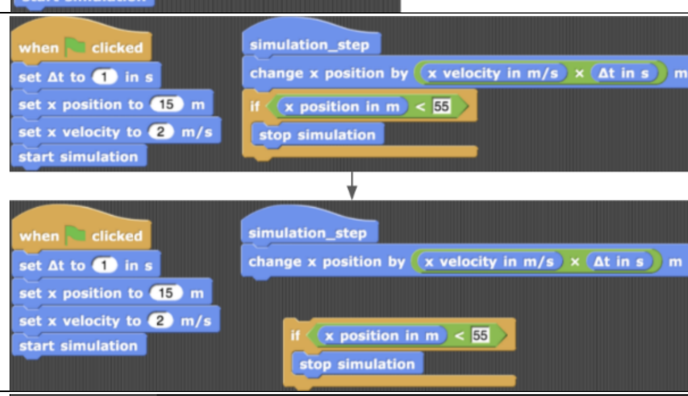
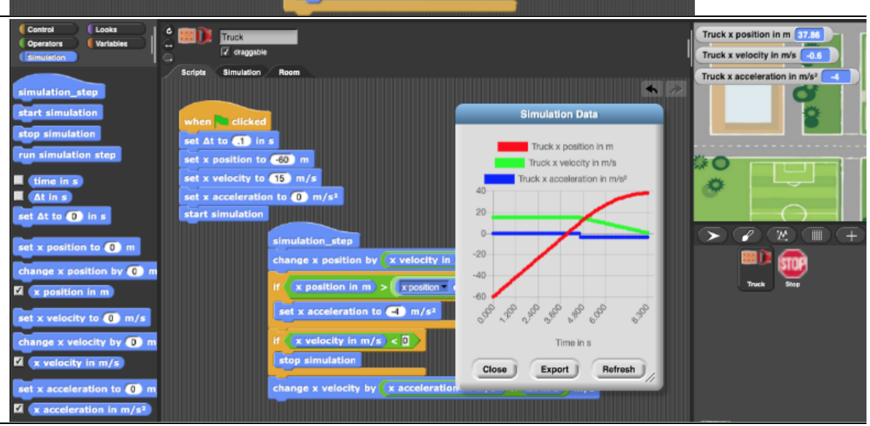


Figure 4. C2STEM task model.

Table 2. Example C2STEM Actions from Task Model

Actions description	System view
REMOVE action: The student removed the "change x position by [expression]" DSML block and it is no longer visible on the script area (where block-based language elements are added to develop the task's executable code).	
SAVE action: In this scenario, the student moved a block (the "if" statement) from the executable code (code connected to a flag that will execute when the green flag is clicked) to the script area. The code is still visible and intact, allowing for later use if the student elects to do so. This action is similar to a student commenting out their code.	
An OPEN action will open the table or graph (depending on selection) and a CLOSE action removes the data tool from view. In order for values to be added to the data tool, a variable must be selected (the RECORD action in the task model). When variables are recorded, the values will also be presented on the stage (indicated by the oval "Truck x position in m [value]" shown on the top right of the image).	

Over the course of three tasks, the log files included (2,205) actions performed by five groups (i.e., 15 sessions). For this study, we track the evolution of the students' computational modelling over three tasks of differing difficulty by looking for differences in their action patterns using the process mining (PM) algorithm (Günther & Van Der Aalst, 2007). Process mining methods reveal the frequencies and sequences of actions by exploring dependencies between actions (van der Aalst, 2011). The analysis of the created log file was performed using ProM 6.9 (2019). We used the heuristic mining algorithm as the dependency values, ranging between –1 and 1, between all possible combinations of events and computed using the following formula (Weijters, van Der Aalst, & De Medeiros 2006, p. 7):

$$A \Longrightarrow_w B = \left(\frac{|a >_w b| - |b >_w a|}{|a >_w b| + |b >_w a| + 1} \right)$$

Based on an event log W, the strength of a dependency relation between two events, A ⟹w B, is computed using the number of times event a is followed by event b, subtracted from the number of times event b is followed by event a, and divided by the number of occurrences of these two relations, plus 1. The number of relevant (a follows b) event sequences and its opposite (b follows a) together influences the dependency value. When building a dependency graph, we set the following thresholds to their default values for events to be modelled (see Weijters et al., 2006 for more details):

1. Dependency measure threshold: minimum strength of dependency between events.
2. Positive observation threshold: minimum value of supporting dependency frequency between events.
3. Relative to best threshold: minimum value of the difference between event dependency value with the maximum dependency value.
4. Length-one threshold: minimum value of same event dependency.
5. Length-two threshold: minimum value of looping pair event dependency.

To check the behaviour proportion on event logs in the model, the fitness significance is calculated (defined in Kurniati, Kusuma, & Wisudawan, 2016; Sonnenberg & Bannert, 2015) using the following formula:

$$\text{fitness}(\sigma) = \frac{1}{2}(1-\frac{m}{c}) + \frac{1}{2}(1-\frac{r}{p})$$

The fitness significance is calculated by replaying every trace using the following four measures: 1) p (produced token), 2) c (consumed token), 3) m (missing token), and 4) r (remaining token). The higher the fitness value, the higher the similarity between the model and the activity sequence. In a review of state-of-the-art process discovery algorithms, De Weerdt et al. (2012) found that the heuristic miner algorithm was especially suitable in a real-life setting, and the algorithm has been successfully used in the past for discovering collaborative regulation processes (e.g., Sonnenberg & Bannert, 2015; Sobocinski et al., 2017).

4.2. Identifying Processes and Types of CPS Regulation During Computational Modelling

All discourse from the video-audiotaped sessions were coded using the software ATLAS.ti 8 and the output of co-occurrence of data was used as input for creating CORDTRA diagrams. The coding procedures focused on students' discourse interaction. In stage one, the macro and micro levels of cognition and metacognition regulation utterances were identified (based on indicators derived from Klahr & Carver, 1988; Winne & Hadwin, 1998). As each annotation can potentially contain several regulation processes, the unit of analysis was a sentence segment, which in most cases was a complete subordinate or dependent clause. The micro level of cognition processes during computational modelling were identified when student talk was about problem identification, problem representation, and/or model test and assessment (Klahr & Carver, 1988). On the other hand, the micro-metacognitive processes were identified when students negotiated and developed task understanding, planning and enactment, and monitoring and evaluation. We also identified off-task talk. Table 3 represents our coding scheme used to identify macro- and micro-cognitive and metacognitive regulation processes with example threads. Inter-rater reliability for the cognitive and metacognitive regulation of process coding was checked for 20% of the data, resulting in Cohen's kappa values of k = 0.94 and 0.89, respectively, which represents excellent agreement (Blackman & Koval, 2000).

Table 3. Coding Scheme for Cognitive and Metacognitive Regulation Schemes Along with Description and Representative Quotes from Transcripts

Macro level	Micro level	Description	Example
Metacognition	Problem understanding	• Analyze and understand the task requirements • Describe how reading about a specific item could help them to solve a problem • Recognize the problem by giving examples	S3: "It will let us accelerate in 2D air because that is when we start factoring in gravity. So then, start simulation, simulation step."
	Planning and enactment	• Determine and describe: − what actions, blocks and resources are needed or what to do next − a goal for the work to be done; the group sets a task-specific goal • Applying appropriate strategy changes	S2: "let's make this an if else and put stop simulation in the else so once it gets there it should stop moving"
	Monitoring and reflection	• Provide feedback to ideas or solution with explanation • Explain and analyze components that are responsible for the misbehaviour • Use data tools to evaluate codes	S1: "You're right. Do you know how many operators it's going to be? It's going to be so many operators"
Cognition	Problem identification	• Read the task instructions but do not interpret the task requirements • Recognize that there is a problem (error) but provide no explanation	S1: "the boat doesn't stop"
	Problem representation	• Externalize actions or steps while generating the model but provide no explanation • Verbalize trial and error or guess and check strategies	S1: "change y position in here. "Just do trial and error"
	Model test and assessment	• Ask for testing their model but do not follow up with debugging • Students do not know what to do after a negative outcome • They change their course of action to avoid a negative outcome again, but they do not look for evidence to know if the change will lead to a positive outcome	S1: "I don't know what's going on?" S3: "Just click green flag"

In the second stage of video coding, types of shared regulation were coded by first identifying segments that were promising as evidence of initiative-response relationships between turns of talk (Kneser et al., 2001). Thus, the coding unit was at the episode level, which means that coding could be assigned to a single talk turn or alternatively to several consecutive talk turns together, depending on the content of the group's interaction. We relied on prior analysis schemes as a guide to develop our collaborative regulation coding categories (Malmberg et al., 2017; Emara et al., 2017). Two coders coded for the types of regulation based on the indicators in Table 4. Inter-rater reliability was checked by calculating Cohen's kappa value, which resulted in very good agreement for SSR (k= 0.81) and SR (k= 0.74).

Table 4. Categories Describing Regulation Types During CPS and their Social Interaction and Transactive Indicators

Regulation code	Social interaction and interactive indicators in CPS processes	n-gram examples	Part of speech (POS tags)
SR	• Students not contributing to each other's work, or any form of communication. • One student externalizes his knowledge or activities but the partner does not contribute to the discussion; may show some signs of joint attention by back-channelling or nonverbal reactions (e.g., nodding, eye contact). • One of the students explains their problem-solving process (thinks aloud) to the others; the others confirm/repeat to externalize the approach but do not add anything new to the discourse. • *CPS Approach Topics:* Instructive, Directive	"I know how," I know what"	Pronoun, Verb 1st person, Wh-adverb (PR, VBP, WRP)
SSR	• Students build on each other's contributions using argumentation, explanation, and consensus on next steps. • Proposals and ideas for next steps are not passively accepted but negotiated and may lead to protracted discussions by adding and integrating new information by combining each other's ideas. • Students adapt by negotiating a different strategic method to the task (debating each question as a group). • Discussions involve what to do next (and why), how a proposed action fits into an overall plan, how to proceed after an assessment, and so on. In these discussions, students are likely to verbalize their problem-solving strategies and demonstrate CPS processes. • *CPS Approach Topics:* Interactive, Explanatory	"Think if you" "Look at this"	Verb 1st person, Subordinate Preposition, Verb 2nd person (VBP, IN, VBZ)

4.3. NLP Analysis

Trace data that does not integrate discourse typically lacks key contextual and problem-solving information to derive a more complete understanding of the group processes. Identifying key phrases as relevant for types of regulation of CPS may help us in developing feedback mechanisms based on the recognition of applied phrases. In this study, we seek to extend our prior human-coding approaches to discourse analysis to better understand how finer-grained aspects of the group regulation in CPS influences computational modelling by utilizing natural language processing (NLP; Rosé et al., 2008), and specifically the text mining approach, Part-of-speech (POS) n-grams (Sullivan & Keith, 2019). In our analysis, we focused on the following features (building on Rosé et al., 2008; Sullivan & Keith, 2019), available in the publicly downloadable version of Rapidminer, to exploit context for developing a more complete understanding of the student discourse.

4.3.1. Data Pre-Processing

Before representing the data using POS n-gram, the data was processed by the removing repeated words (e.g., wait wait wait), strings with repeating characters (e.g., Noooo or ohhhh), and punctuation. All data was made lowercase. An n-gram word-based tokenizer was created to slice the text based on the length of n. After tokenizing the data, we transformed the tokens into a standard form (i.e., stemming). Stemming changes the words into their root, and decreases the number of word types or classes in the data in order to allow some forms of generalization across lexical items, for example the words stable, stability, and stabilization all have the same lexical root (Rosé et al., 2008).

4.3.2. POS n-grams

We then identified n-gram word sequences of text that occurred more often than expected. To perform the n-gram analysis, the transcripts of the words uttered by students were broken down into uni-, bi- and/or tri-gram word segments. We combined n-gram with POS n-grams. POS tagging tokenizes individual words and then utilizes computational methods to assign a POS (such as noun, verb, coordinating conjunction) to each word. The choice of n-grams based on POS tag patterns has been recommended in recent studies (e.g., Sullivan & Keith, 2019) because it helps not only to focus on the word sequences but also on their grammatical categories. They can be used as proxies for characterization of syntactic structure (Rosé et al., 2008). Thus, they may capture some formal information, such as the difference between "the answer, delta t is" vs. "delta t is the answer."

In this study, we focused on tri-grams because we found that the tri-gram unit of analysis helped us to better understand macro- and micro-regulation processes as well as how the different types of regulation are activated in the CPS discourse while building computational models. This approach is like that of other researchers (cf. Bakliwal et al., 2011; Sullivan & Keith, 2019). According to our data, we considered a phrase of "not know why" to demonstrate how uni-gram and bi-gram do not carry sufficient information for understanding students' regulation discourse. For example, when we move to bi-gram as a unit of analysis the results are "not know" and "know why." The bi-gram "know why" has a sentiment towards positive feeling of knowledge and "not know" is negating the feeling of knowledge. The tri-gram "not know why" gives enough information to classify the tri-gram in the negative context. For the purposes of this analysis, we selected more frequently occurring tri-gram segments for each group, deemed unique tri-grams. A unique tri-gram is defined to have occurred in a minimum of five segments of text across all sessions.

Following the work of Sullivan & Keith (2019), we combined our manual coding of regulation of CPS with automated suggestions derived from n-grams and POS matching. POS taggers helped us identify types of utterances that could be mapped onto the target regulation process, utilized in Malmberg et al. (2017) and Emara et al. (2017), and defined in the "Social interaction and interactive indicators in CPS processes" column of Table 4. Using domain expertise and the theoretical model of CPS regulation, we mapped each POS tag string to a code in the regulation of CPS coding scheme (Tables 3 and 4). By using POS n-gram, it was not the specific words that mattered, rather it was the grammatical role the words played in the overall structure of the utterance that mattered. The mapping of POS n-grams to CPS regulation codes included manual evaluation of the n-grams in the context of the computational modelling currently being performed. Example n-grams and POS tags from this process are provided in Table 4.

4.3.3. Feature Extraction

We applied feature extraction using TF-IDF to reduce the text feature size and avoid analysis in high dimensional feature spaces. **TF-IDF** is a weighting metric often used in information retrieval and natural language processing. It is a statistical metric used to measure how important a term is to a document in a dataset, in our case, the collaborative discourse recorded during each task. The goal of this application was to differentiate unique tri-gram use across three different computational modelling tasks (each modelling task having a unique document, or group conversation) to answer RQ3. A term importance increases with the number of times a word appears in the document; however, this is counteracted by the frequency of the word in the corpus. One of the main characteristics of IDF is that it weights down the term frequency while scaling up the rare ones (Mohammed & Omar, 2020). For example, words such as "the" and "then" often appear in the text, and if we only use TF, terms such as these will dominate the frequency count. However, using IDF scales down the impact of these terms. Resulting terms from the POS tri-gram analysis were ranked with this measure. Three terms with the highest percentage from each task's conversations were identified (Figure 8b, 9b, 10b) and manually compared to our researcher generated CPS approach topics (Table 4) in order to evaluate group CPS based on our hypotheses described in Table 1.

4.4. CORDTRA Diagrams

To understand in more depth how regulation macro- and micro-level processes unfold over time and how these processes emerge according to types of regulation, we present a case example for each task using the Chronologically Ordered Representations of Discourse and Tool-Related Activity (CORDTRA) diagrams (Hmelo-Silver et al., 2008). We used this to study distinct groups' actions for the three different tasks in the C2STEM computational modelling environment.

4.5. Scoring Model-Building

Model-building tasks were scored using the rubric outlined in Table 5. The students' model building scores were normalized to a [0,1] value using the measure, -$score - min/max - min$).

Table 5. Rubric for Evaluating Student Models for the Challenge (C) Tasks and Instructional (I) Task

	Description/Example	Points		
Expressing physics relations in a computational model		**1D (C)**	**2D (I)**	**2D (C)**
Program initializes (x/y) position to the correct starting value	`set x position to (0) m` the variable must be set to the initial value stated in the problem	1	1	1
Program initializes (x/y) velocity to the correct starting value	`set x velocity to (2) m/s` the variable must be set to the initial value stated in the problem	1	1	1
Program initializes (x/y) acceleration to the correct starting value	the variable must be set to the initial value stated in the problem (e.g., for y-direction): `set y acceleration to (gravity in m/s²) m/s²` Or hardcoded -9.8; must update velocity correctly (above)	1	-	-
Program initializing heading to the correct starting value	`set heading to (atan of ((4) / (5))) deg` the variable must be set to the correct value	-	1	1
Program expresses correct relations among velocity, position, and time, and correct units for each	`change x position by ((x velocity in m/s) × (Δt in s)) m` must use equations or use multiply operator and hardcode the values for velocity and delta t; partial credit for hardcoding change (e.g., if velocity is 2 and delta t is 1, but they use the change block and insert the value "2" without using operators/expressions	1	1	1
Program expresses correct relations among acceleration, velocity, and time, and correct units for each	`change x velocity by ((x acceleration in m/s²) × (Δt in s)) m/s` see similar description for changing position	1	-	-
Program expresses correct values for updating velocity (2D) or acceleration (1D A)	the associated variable (based on task description) is updated to reflect the changing physics behaviour of the model (e.g., acceleration is reset to simulate the slowing down of the truck)	1	-	1
Program updates heading correctly	`set heading to (atan of ((5) / (2))) deg` update the variable with the correct value	-	1	1
Program accuracy	The model achieves the desired simulation goals (e.g., completes the physics behaviour) described in the task instructions in order to solve the assigned problem.	1	1	1
Using programming concepts to model physics phenomena				
Program makes the distinction between actions that need to happen once during initialization and actions that need to be repeated in the simulation step	Set actions happen under green flag (exception: set acceleration in the truck stop task); change typically happens under sim flag; this is the only rubric in which the "simulation step" structure is scored	1	1	1
Program initializes variables utilized in updating the simulation behaviour	If a variable is updated and/or used in the updating of the object's behaviour, it needs to be initialized; no extraneous variables set	1	1	1
Program initializes delta t for use in modelling desired relationships (see STEM domain)	Distinguishing delta t from the above rubric item; potential for correlating initialization of delta t to understanding of dynamic behaviour changes	1	1	1
Program sets initialized variables in the correct fashion	Set block used (cannot initialize variable by using change block); `set y acceleration to (gravity in m/s²) m/s²` generalizability	-	1	1

Table 5. *(Continued).* Rubric for Evaluating Student Models for the Challenge (C) Tasks and Instructional (I) Task

	Description/Example	Points		
Using programming concepts to model physics phenomena		**1D (C)**	**2D (I)**	**2D (C)**
Program updates variables with correct function	Change blocks used to update variables in the correct manner, if new value needs to be set (e.g., acceleration changes) then set block used	1	1	1
Program updates variables with correct operators/ expressions	Scoring use of operators and expressions to create generalizable code; no hardcoding if possible	1	1	1
Program updates initialized variables in the correct sequence	Change velocity before changing the position at each simulation step	1	-	1
Program updates/ sets initialized variables under the correct conditions	For instance, the truck stop handles significant conditional logic regarding speed limits; look ahead distance to start slowing down to a stop	2	-	2
DRY (Don't Repeat Yourself) principle achieved	No duplicate code; all connected code is reachable/can be executed	1	1	1
Simulation ends based on stopping logic	Correct condition set to stop simulation appropriately	1	1	1

5. Results and Discussion

5.1. RQ1: What is the nature of CPS regulation activated by students when they work in groups on three different types of physics modelling tasks of varying complexity?

Table 6 depicts increasing adoption of SSR regulation by the five groups as the task complexity increases. Concurrently, self-regulation decreases commensurately from the instructional task to the 2D challenge task. Similarly, adoption of metacognitive planning grows from the 2D instructional task (5%) to the 2D challenge task problem (14%). Table 6 shows a dominance of model test and assessment and monitoring and reflection in all tasks. Despite an increased adoption of model testing and assessment in the 2D instructional and the 1D challenge tasks, this evolution is less pronounced compared to the trends in metacognitive monitoring for the 2D challenge problem (31%). In contrast, metacognitive task understanding remains small and rather stable. This echoes the findings of Iiskala et al. (2011) since they detected more and longer episodes of SSR for difficult problems as compared to moderately difficult and easy problems.

Table 6. Mean Frequency of Occurrence of Cognition and Metacognitive Regulation During the Three Tasks of Different Complexity (shown as percentages, with standard deviation in parentheses)

Modules	Metacognition			Cognition			Regulation types	
	Task understanding	Planning and enactment	Monitoring and reflection	Problem identification	Problem representation	Model test and assessment	SSR	SR
Instructional Task: 2D constant velocity (n= 5 groups)	2% (1%)	5% (1%)	20% (4%)	4% (2%)	30% (7%)	39% (8%)	39% (3%)	61% (9%)
Challenge Task #1: 1D motion with acceleration (n= 5 groups)	6% (1%)	7% (2%)	21% (4%)	9% (2%)	23% (2%)	34% (3%)	47% (9%)	53% (7%)
Challenge Task #2: 2D constant velocity challenge (n= 5 groups)	5% (4%)	14% (1%)	31% (2%)	7% (2%)	18% (5%)	25% (3%)	68% (4%)	32% (2%)

5.2. RQ2: How do students' self- and shared regulatory activities correlate with their performance (model-building scores)?

To address this question, the partial correlation coefficient between computational modelling scores (Table 7) and SSR proportion of problem solving per group was calculated (Table 8). When we control for proportion of total time spent, number of actions in each group, a positive relationship between the frequency of SSR and the model-building scores ($r = 0.30$, $p=0.29$) was found. Specifically, there was a significant positive relationship between the frequency of SSR and the physics score ($r= 0.68$, $p=0.007$). Similar results were also reported by Oshima et al. (2020) and our results confirm other findings on the important relations between collaborative metacognitive regulation (Winters & Alexander, 2011) and socially shared regulation (Järvelä et al., 2016) and learning performance. A plausible explanation is that students find it hard to solve the challenge tasks on their own, and this increases their awareness and need for socially shared regulation. This echoes Sobocinski et al. (2017) who found that when groups are confronted with challenges, they adapt their strategies and explore different options in their collaboration by jointly regulating their problem-solving activities.

Table 7. Group Performance on Model-building Scores and Normalized Scores

Group Performance	Instructional Task: 2D constant velocity problem			Challenge Task #1: 1D motion with acceleration problem			Challenge Task #2: 2D constant velocity challenge problem		
	Total score (13)	PHY score (5)	CT score (8)	Total score (17)	PHY score (6)	CT score (11)	Total score (18)	PHY score (7)	CT score (11)
Mean Score (SD)	11.8 (1.4)	4.2 (0.7)	7.6 (0.8)	15.2 (1.3)	5.2 (0.4)	10 (1)	14.8 (2.4)	5 (1.2)	9.8 (1.3)
G1	9.5	3.5	6	14	5	9	10.5	3	7.5
G2	12	4	8	16	5	11	16.5	6	10.5
G3	13	5	8	17	6	11	15.5	5	10.5
G4	11.5	3.5	8	15	5	10	16	5	11
G5	13	5	8	14	5	9	15.5	6	9.5

Table 8. Correlation between Group Performance on Model-building Scores, and Two Types of Regulation (SSR and SR)

Type of regulation		Physics score	CT score	Total scores
SSR	Correlation	.688*	.168	.300
	Sig. (2-tailed)	.007	.565	.297
SR	Correlation	-.601**	-.163	-.291
	Sig. (2-tailed)	.005	.566	.301

To illustrate these findings in more depth, we describe a case example demonstrating the change in a group's way of thinking and talking about how to solve the problem supported by CORDTRA diagrams (Figure 5, 6, 7). We selected Group 4, who performed at or above the median split of the computational modelling task scores on all tasks (see Table 7).

Group 4's instructional task performance (Figure 5) demonstrated higher frequency of SR than SSR, with a focus on cognitive processes, especially at the beginning of the task (when the focus on scaffolding is greater). Transitioning to the first challenge task, the group switched between SR and SSR episodes throughout the model construction process, but as can be seen in Figure 6, the frequency of SSR increased by the end of task. We hypothesize that this may be a result of increased debugging based on our previous findings targeting collaborative regulation in debugging (Emara et al., 2020). Finally, as shown in an example in Figure 7, Group 4's regulatory processes improved by the time they worked on the final 2D challenge task. The group began the modelling process enacting socially shared regulation with metacognitive planning and task analyses followed by plan monitoring. This decreased the conflicts in problem interpretation and increased students' shared task understanding (see examples in Table 11). This is supported by Sobocinski et al. (2020) and Paans et al. (2019) who reported that when students start their task by taking turns and combine planning with monitoring and reflection while simultaneously giving feedback to each other on strategies to employ in developing the problem solutions, this can lead to SSR (Sobocinski et al., 2020) and productive interaction (Paans et al., 2019). These results indicate the development of improved regulation processes over time during computational modelling tasks.

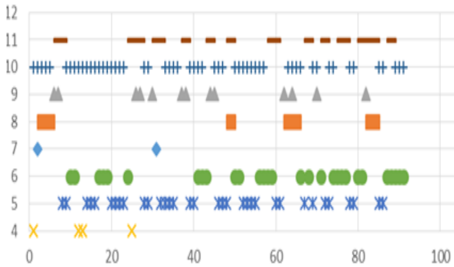


Figure 5. CORDTRA diagram: sequence of regulation processes and types over time in Group 4 at the instructional 2D task

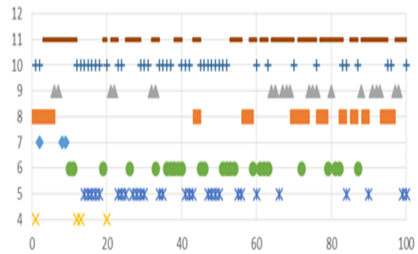


Figure 6. CORDTRA diagram: Sequence of regulation processes and types over time in Group 4 at the 1D challenge task)

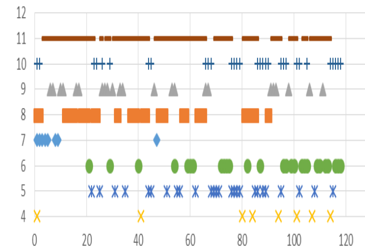
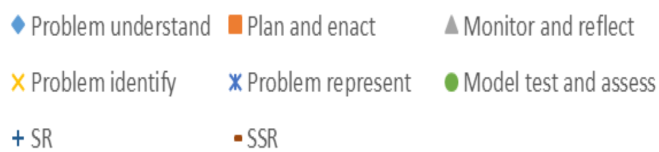


Figure 7. CORDTRA diagram: sequence of regulation processes and types over time in Group 4 at the 2D challenge task



5.3. RQ3: How do action patterns and problem-solving strategies derived using NLP emerge across collaborative computational modelling tasks of varying difficulty?

To understand how student regulation of CPS processes affects their computational modelling strategies, we first applied a process mining (PM) technique to compare group sequential action patterns across the three different tasks of varying difficulty. We also extended our interpretation of the process models with the derived POS n-grams. We provide three case examples (one from each task) to illustrate these findings in more depth.

Leveraging the analysis method described in Section 4.1, the PM results are presented in Figures 8a, 9a, and 10a. The numbers in the boxes (nodes) represent the frequencies of the actions, the arcs display the ordering of actions and the number shows the strength of the sequential relation between the two actions. The fitness values for the process models are 0.69 for the instructional task, 0.54 for the first challenge task, and 0.62 the second challenge task. The process models for tasks 1, 2, and 3 (Figures 8a, 9a, 10a) share one similarity. All three models contain a one-directional path from CONNECT to RECORD. Transitions initiating from PLAY occurred frequently, reflecting the use of observing the simulation behaviour as a frequent checking mechanism during model construction and debugging.

The NLP analysis produced 35,316 unique POS n-gram segments of text (i.e., uni-grams, bi-grams, and tri-grams). As expected, we also see a sharp decline in POS n-gram frequencies as n increases, with 4003 tri-grams. After completing the POS tagging and feature extraction, a close analysis of the resulting ranked POS tri-gram segments differentiated by task showed a clear evolution in student regulation processes and approaches as they addressed tasks with increasing computational modelling difficulties in the C2STEM environment. We present the three unique POS tri-grams uttered during each task in Figures 8b, 9b, and 10b. A visual inspection of these bar charts shows how the groups change their way of thinking and talking about how to solve the problem, and specifically how to use the C2STEM environment tools to help solve the problem.

The POS tagging of the tri-grams allows us to easily associate one word either with reasoning or argumentation in terms of the types of regulation codes in Table 4. With similar findings reported by Sullivan & Keith (2019) and Lobczowski et al. (2020). The first, second or third person "I, you, we" may refer to how students share with the group members in the problem context; the words "but" and "wait" may indicate students co-constructing knowledge through challenging; the word "and" indicates elaboration; the words "because" and "if" show evidence or hypothesis generation as well as words that identify an entity; and "it" may refer to any of the data tools in the C2STEM environment. In our analysis, for example, student discourse commonly included "I_know_how" in the instructional task (Figure 8b). The use of "I" without accompanying terms correlated with co-construction are indicative of SR, while the unique POS tri-gram "think_if_you" from the final challenge task (Figure 10b) indicates SSR through the proposal of a new idea to another group member.

Given the differences in task complexity between the three modelling tasks, the result of PM and NLP techniques will be examined for each task in the following sections.

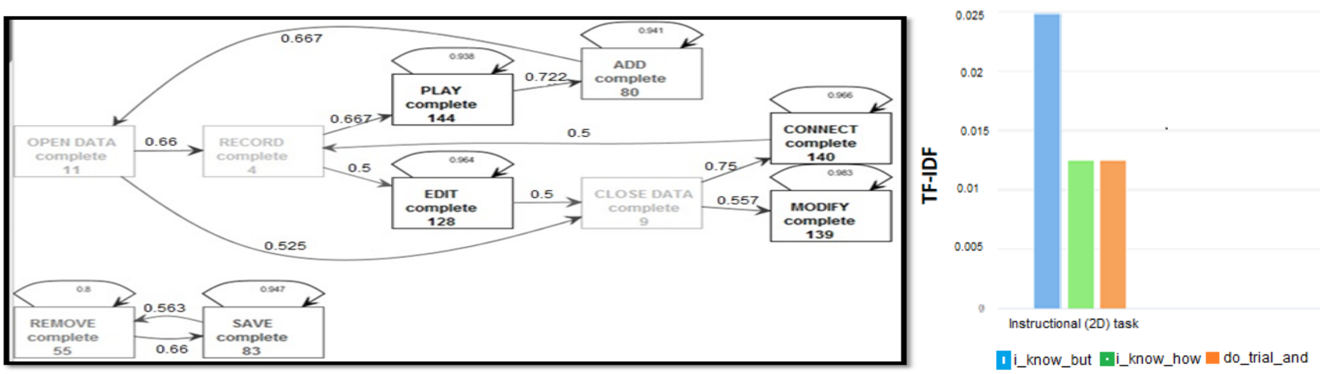


Figure 8. a) Process model of groups during the instructional 2D task;
b) bar chart of the most frequent three tri-grams during the instructional 2D task.

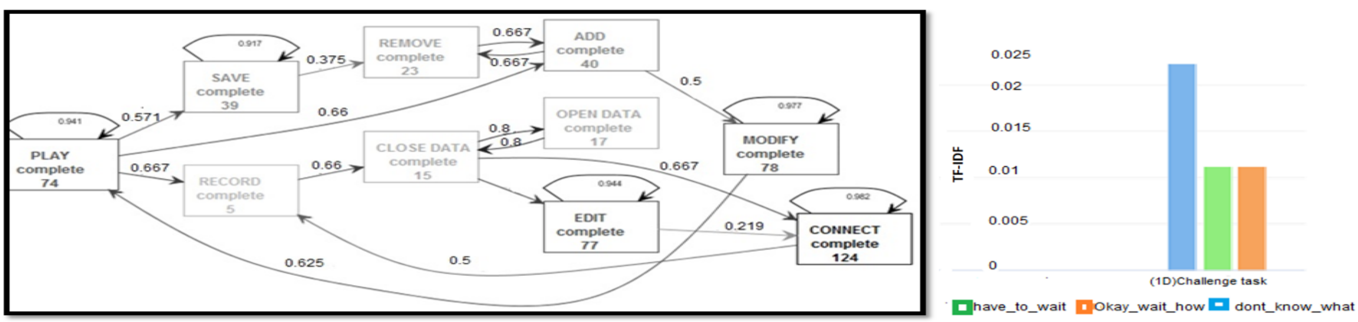


Figure 9. a) Process model of groups during the 1D challenge task;
b) bar chart of the most frequent three tri-grams during the 1D challenge task.

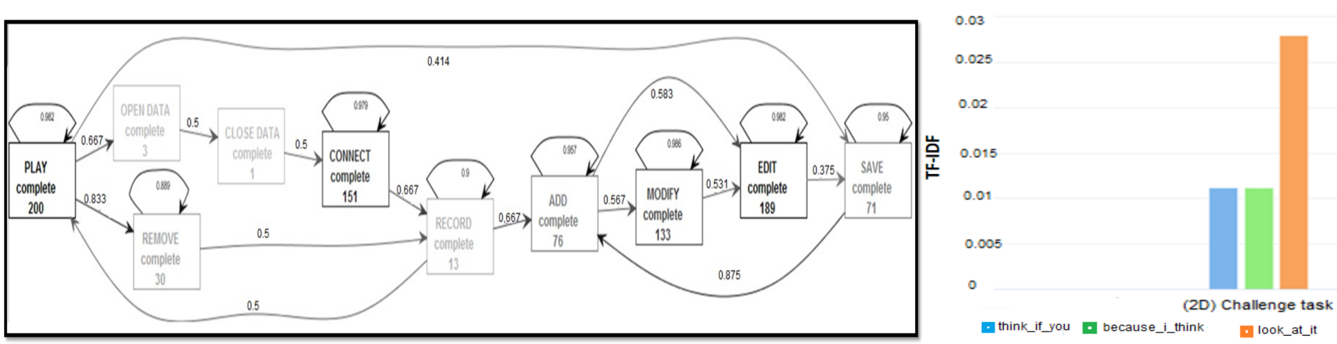


Figure 10. a) Process model of groups during the 2D challenge task;
b) bar chart of the most frequent three tri-grams during the 2D challenge task.

5.3.1. PM and NLP Analyses of the Instructional 2D Task

A key observation in the process model of the 2D instructional task (Figure 8a) is the position of REMOVE actions. The increase in the frequency of eliminating items from the executable model clearly resulted in an isolated process between “REMOVE” and “SAVE” activities in the process model (Figure 8a). “REMOVE” actions involve the removal of code. The deleted code is no longer viewable. This is separate from “SAVE” actions where students disconnected the potentially erroneous code from the model, but left it in the “Script” area, hoping to make corrections later and reintroduce the block(s) into the model. The fact that “REMOVE” and “SAVE” activities, which may be considered useful for debugging, were not connected to the rest of the problem-solving actions implies that these actions were not combined with other actions in a systematic way. In this model, the “PLAY” actions seem to limit themselves to testing a single construct of their model. As Figure 8a shows, the actions preceding and following “PLAY” are mostly to/from the “ADD” action. This could include the adding of a block and “PLAY” of that block to evaluate its impact on the simulation behaviour. This is separate from modifications to the existing code (through actions such as “CONNECT,” “MODIFY,” or “EDIT”). These results indicate that students may not have been confident in their identification of model limitations or errors as they elected to add potential code updates to the stage, instead of modifying existing, executable code (e.g., code connected to the "Green Flag").

The POS tri-gram analysis indicates the groups in the instructional task worked individually more often than in the challenge tasks. The three tri-grams in the Figure 8b indicate that the students used a trial-and-error approach such as "do_trial_and" or individual discourse with a single person pronoun "I_know_but" or "I_know_how." These findings are supported by previous work (e.g., Barron, 2003; Dowell et al., 2020; Malmberg et al., 2017; Roschelle & Teasley, 1995) where the authors confirmed that simply placing participants in groups does not guarantee successful collaboration strategies. Further, previous findings show that unproductive interactions (see examples in Table 9) decrease problem-solving success (Chiu & Khoo, 2003).

5.3.2. PM and NLP Analyses of the (1D) Challenge Task (Truck Problem)

The PM of the groups during the first 1D challenge problem (Figure 9a) shows that "OPEN-DATA" is bi-directionally connected to "CLOSE-DATA," which in turn is connected to "CONNECT." "REMOVE" is also bi-directionally connected with "ADD," which in turn is connected to "MODIFY" then "PLAY." The increase in actions following PLAY compared to the instructional task may indicate better usage of the visualization as a tool to support model adjustments (including collaborative discussion of the resulting visualization). For instance, groups increase the combination of PLAY to SAVE actions, saving executable code for potential later use, and PLAY to REMOVE, removing code from scripts, indicating processes involving the identification and correction of code errors. These regulation loops generalize to multiple bi-directional cycles between "REMOVE" and "ADD" blocks activities that groups engaged in before "MODIFY" activities.

Linguistically, the three POS tri-grams in Figure 9b indicate uncertainty, e.g., "don't_know_what," or being cautious "have_to_wait," "okay_wait_how," which were more common among groups at the first 1D motion challenge problem (i.e., working without scaffolds) than among groups in other tasks. These initial results confirm a result previously presented by Worsley & Blikstein (2011), which indicates that realization of lack of expertise tends to decrease learner confidence. In terms of regulation of CPS, groups tried to understand the physics problems in the first challenge task (without scaffolds) by monitoring the visualization and carefully planning to build their model as demonstrated in their discourse in Table 10.

5.3.3. PM and NLP Analyses of the 2D Challenge Task (Final Boat Problem)

Regarding the evolution of CPS strategies from the 1D challenge task to the 2D challenge task (the most challenging of the three tasks), groups engaged more in socially shared regulation (SSR) to jointly build their computational model. This yielded CPS strategies that used the graphing or table data tools as part of a Code Assessment strategy (shown in their process model in Figure 10a). The "OPEN DATA" and "CLOSE DATA" actions involve the clicking of a button to view or close the graph or table tool. Moreover, transitions from "CONNECT" to "RECORD," and from "REMOVE" to "RECORD," are observed several times. Interestingly, this observation indicates that these regulation loops occurred more frequently during the high challenge task, although the frequencies of cognitive problem representation and model test decreased from the instructional task to the 2D high challenge task as discussed above in RQ1. An explanation could be that the groups improved their metacognition discussion of CPS, possibly because they adapted to the demands of the task. Therefore, when students engage in SSR, the process cycles mostly result in further activities for both testing and debugging (Figure 10a). It appears that the activities following the PLAY (of the simulation) are targeted at adaptation of models to address problems encountered in their previous run of the simulation. The process map for the SSR groups seems to suggest that students in the high challenge task (Task #3) put a lot more effort into the validation of behavioural aspects of their models by means of code assessment tools (Figure 10a).

Concerning the NLP POS tri-gram analysis, Figure 10b provides three tri-grams of the groups engaged in more SSR. The first pattern "think_if_you" and second pattern "because_i_think" demonstrate that the groups worked together to collectively build and evaluate part of their model (see example 1 in Table 11). Distinct from the previous tasks, the use of "because" indicates that the groups were now more likely to provide arguments in support of their suggestions. This may be supported by the increased usage of data tools to identify errors or features of the model and use the information to explain model-building choices to other group members. Another interesting unique word pattern resulting from n-gram text mining is "look_at_it," which could be used as a marker for joint attention when they jointly share regulation of their CPS in a synchronous environment. These findings are supported by previous work by Bangerter (2004) and Schneider et al. (2018) where authors highlighted the importance of joint attention in small groups of students. More generally, there is a large body of evidence showing that joint attention is a central mechanism for effective collaboration (e.g., the process of building a shared understanding has been extensively studied by psycholinguists under the name of grounding; Clark & Brennan, 1991).

5.4. Applying Our Multimodal Learning Analytics (MMLA) Framework

To illustrate the findings in more depth, we describe case examples that illustrate changes in group CPS processes by applying our MMLA framework. The groups selected include Group 4 (described in Section 5.2) and Group 5, who performed below the median split of task scores on the instructional task and improved to the 2nd highest overall score and highest CT score by the final challenge task. The talk examples in Tables 9–11 were identified based on the result of tri-gram analysis.

5.4.1. Case Example of the Instructional 2D Task

The examples in Table 9 demonstrate initial usage of SR during the instructional task (the bolded text identifies the POS tri-gram from Section 5.3). Group 4 initiated their computational modelling process with predominantly SR codes (as shown in Group 4's CORDTRA graph in Figure 5), but concluded with SSR, as seen in Table 9, Example 2. While there was instructive discourse, initial attempts were made to explain reasoning, although the explanations are domain-focused and do not indicate use of system tools (e.g., data tools). In this case, the S1 could be identified as the Driver (Dowell et al., 2019) or Owner (Pino-Pasternak et al., 2018) when students were investing a high degree of effort in collaborative discussion and displayed self-regulatory and social-regulatory skills. However, by the end of the session, there was a switch to more SSR, as they noticed they were making several errors. So, they applied metacognitive planning and monitoring to figure out why their debugging actions were not working, and how to fix their errors.

Group 5 demonstrated more SR throughout and the use of "do trial and error" indicate difficulties in systematically planning and building their model. This coincides with the process mining results described above. An interesting finding in this discourse is the lack of reaction to group members' contributions (for instance, the response to the trial and error suggestion). The implication is that the contribution was not seen as valuable by the other group members. This constitutes an ignored socially shared regulation attempt (e.g., Molenaar, Chiu, Sleegers, & van Boxtel, 2011). This pattern of talk was identified by Dowell et al. (2019) as *Socially Detached:* when the pattern appears to capture students who were not productively engaged with their collaborating peers, but instead focused solely on themselves and their own narrative. A possible explanation for this is that when the students are first introduced with new domain content during the scaffolded instructional task, students with higher prior knowledge of the content may assume control, resulting in greater self-regulatory and cognitive regulation processes (e.g., instructing on next steps with no explanation, as seen in the Group 5 example). This may pose a problem as computational models require the translation of domain knowledge into computational form.

Table 9. Groups 4 and 5 Discourse: Regulation Processes and Types During 2D Instructional Task

Student "words" and [actions]	Regulation processes	Regulation types
Example 1 — Group 5		
S1: [first moves the change x velocity block, but then removes it] S1: "…or we would want it to be … yeah change velocity by … so" S1: [clicks on a river object to show pre-programmed velocity.] S1: [clicks on boat object to return to code].	Problem representation	**This talk identified as self-regulation (SR)** S1 is externalizing his thinking during doing actions, and in doing so, involving the partners. S2 is engaged, representing what is needed, however S1 is ignoring S2's proposal while performing all the actions and most of the talk.
S1: I think we should ***do trial and error*** just, uh	Problem representation	
S2: we need to make it go this way	Problem representation	
S1: "Just like experiment with degrees. I mean just like with random numbers show that it faces like that." "Just put random stuff in."	Model test and assess	
Example 2 — Group 4		**Transition of Regulation: SSR/SR demonstrates the types of talk at the end of session** Students activate socially shared regulation as they notice that S1 is making errors. So, they engaged in asking for modifying by S2 then metacognitive plan and monitoring by S1 about how to debug their code. S3 activated self-regulation by asking questions to understand S1 actions without adding contribution.
S1: [Runs simulation to test his model]		
S2: that should be 38 degrees also	Problem representation	
S1: yeah, ***I know but*** we have to find the velocity of this line	Plan	
S1: so square root. square root of 5 over 2, right?	Monitoring	
S3: yes. Why are you using multiplication?	Monitoring	
S1: cause the square	Problem representation	

5.4.2 Case Example of the 1D Challenge Task

Table 10 presents the regulatory processes of Groups 4 and 5 in the first challenge task (1D challenge task). Analysis of the groups' discourse indicates increased use of planning and model testing and evaluation (as shown in Group 4's CORDTRA graph in Figure 6), which may have supported more systematic model construction and debugging. This is further supported by combining the PM and NLP analysis with sequences such as "MODIFY" to "PLAY" demonstrating increased testing and increased use of the data tools, potentially indicating the acquisition of more information from the model to plan.

In general, the results further support the hypothesis that there was adoption of more SSR when students collaborated with each other on a more open-ended task than when they were provided with scaffolds (in the instructional task). These results reflect those of De Backer et al. (2015), who also found that evolution in socially shared metacognition regulation significantly increases in a reciprocal peer collaboration, and significantly decreases when students were prompted by their teacher. Another interesting finding is the higher frequency of POS tri-grams in the first challenge task including the words "wait" and "what" could be consistent with that of Sobocinski et al. (2020), who explained that group members engage in SR at the beginning of a challenge task such as an exam because this may give them the opportunity to spend more of their attention on engaging with the learning material before engaging in SSR, which is presented by our case example.

Table 10. Groups 4 and 5 Discourse: Regulation Processes and Types During 1D Challenge Task

Student "words" and [actions]	Regulation processes	Regulation types
Example 1 — Group 5		
S1: Slow down (reading instructions)	Problem representation	**Transition of Regulation: SSR** At the start of the session, S1 is regulating his modelling by using a cognitive strategy first (read task instruction out aloud) to help in identifying the next step. So, S1 first moves the change x velocity block without explaining his behaviour to others, but then removes it. However, before S1 tries to add a new block, S2 gives instruction to group members by waiting "have to wait" to think and monitor how to organize their next step for the group's code construction.
S1: so then we have to make it go [save then remove x velocity] …	Problem representation	
S2: we ***have to wait.*** so we need to find out once it decreases how long it takes to decrease so we need to add a … wait	Planning	
Example 2 — Group 4		**Transition of Regulation: SSR** By the end, groups check their behaviour by testing their variables and monitoring their approach. S3 was responsible for constructing a model to help the group solve their problem and inviting them to add to or monitor their planning enactment. When S2 has difficulty contributing information about monitoring and understanding the model misbehaviour, S1 successfully prompts S3 to recall their prior experience in the environment so he can bring that knowledge or expertise to the group discussion.
S3: set x velocity to 0 and then it will accelerate	Planning	
S2: I just ***don't know what's*** going on, I don't know but I'm really trying though	Model test and assessment	
S3: okay so	Model test and assessment	
S1: we have to change this, change this, like the sloth remember? Right?	Planning	

5.4.3 Case Example of the (2D) Challenge Task (Final Boat Problem)

Groups 4 and 5 performed higher than the median split of the groups on the final challenge task. Coinciding with our findings comparing learning performance and regulation, both groups demonstrated high usage of SSR (as shown in Group 4's CORDTRA graph in Figure 7). Group 5's discourse is highlighted by the usage of "I think if you" and "because," indicating the sharing of contributions and abilities to reason and explain based on model features. Group 4 added to this approach by using the data tools, providing an example of our process mining and NLP findings in Section 5.3. Through discourse and logged actions, we are provided with a comprehensive understanding of the regulation process and associated environment actions that supported CPS leading to high learning performance.

Table 11. Groups 4 and 5 Discourse: Regulation Processes and Types During 2D Challenge Task

Student "words" and [actions]	Regulation processes	Regulation types
Example 1 — Group 5		**Transition of Regulation: SSR/SSR/SSR** S2 is checking his next step before doing it. S1 does not totally agree but immediately they jointly build on each other's contributions using argumentation, explanation, and coming to consensus on next steps. Group also verbalizes their problem-solving strategies and demonstrates CPS processes.
S2: Velocity of 5?	Planning	
S1: I don't think it's possible to have a velocity of 5 that can go this way. I ***think if you*** had a velocity of…	Monitoring	
S2: I mean…	Problem representation	
S1: unless, ***because I think if you*** angled it really far this way and then as it goes up and it'll kind of go and the river push this back some	Monitoring	
S2: I mean you could have just a really low um, a really low y velocity and a really high x I guess as high as possible	Planning	
S1: so we need to find a way to find x and y values that equal that will equal a hypotenuse of 5, aside from 3 and 4	Planning	
S3: like -5 and 10 with x?	Monitoring	
Example 2 — Group 4		**Transition of Regulation: SSR/SSR/SSR** Here, the group is: 1. Monitoring and reflecting on each other's work 2. Taking appropriate action to solve the problem — enacting plans together 3. Asking questions and responding Overall: S2 is monitoring S1's actions and calls out on a potential error made that S1 subsequently fixes.
S1: how do we...? wait how do we ***look at it***, the variables? [opens data tools]	Monitoring	
S2: I think it's at 0,0	Monitoring	
S1: what did you do last time?	Monitoring	
S3: go to here and display x position, y position.	Planning	
S2: "for the time being we will have it repeat it until get an and operator two equals operators and then x position and y position of this one"	Planning	

6. Conclusions

The aim of this study was to gain insights into regulation of CPS processes — and their evolution over tasks of differing complexity — during computational modelling in science. We applied multimodal analytics (Figure 3) that combined student discourse and activity data from the computer-based model building environment, C2STEM for this analysis. This provided a novel approach, which allowed us to interpret student model building actions by linking them to their discourse. Analyzing the discourse using NLP methods also provided a mechanism for studying the dynamic and evolving regulation processes among collaborating students, and how the different forms of these regulation processes could be linked to their model building performance. Key findings from our analyses demonstrate that:

- Student CPS strategies evolve over time as they work together through computational modelling tasks of varying complexity. This can be seen through increased applications of SSR and, correspondingly, the use of more systematic model-building action sequences (e.g., testing modifications to the code by running the simulation).
- The combination of analytic approaches supports a deeper understanding of the problem-solving process. For instance, increased use of the data tools coincided with increased planning and testing. In addition, an analysis of tri-grams in student discourse showed use of terms like "because" to explain model-building decisions as students collaborated and engaged in socially shared regulation.
- Task scaffolding may impact the regulatory behaviour of groups, with more difficult tasks prompting increased applications of SSR.

Our findings are well aligned with the literature on collaborative learning. For instance, more productive SSR has been shown to be correlated with more productive strategy use, and reflected in the higher model building scores. Recently, Bakhtiar & Hadwin (2020) have shown that the more frequently students engage in SSR, the more frequently they use strategic responses, resulting in more learning and co-construction of knowledge. Moreover, when students engage in more open-ended challenge tasks, such as in challenge task 2, they use more SSR and productive collaboration, using persuasive and explanatory words such as "because" and phrases of higher verbal complexity, such as active requests for sharing contributions (e.g., "think_if_you"). In other words, groups shifted to using more metacognitive collaborative processes rather than focus on lower-level cognition processes. This was illustrated by their greater elaboration of the learning material, integration and synthesis of one another's perspectives, and ideas to make sense of the learning task jointly (Schoor & Bannert, 2012; Weinberger & Fischer, 2006).

The combination of PM, NLP, and CORDTRA findings provide a useful framework for constructing argument prompts to help learners engage in persuasive argumentation, which is also a key science practice that learners must develop. This has been extensively studied in analyzing rhetorical arguments such as claims, grounds, and warrants (see Toulmin, 2001; Verheij, 2005). For example, in our NLP analysis, "Think_if_you," "look_at_it," and "because_I_think" are examples of a claim, ground, and warrant. For the second 2D challenge task, students invoked monitoring processes more often than in other tasks. This was also followed using SSR for evaluation and planning activities. This pattern seems to be more in line with what is referred to as interactive feedback and conflict, where group members seriously and critically consider each other's ideas, and differences in opinion can be resolved (Van den Bossche et al., 2011). These findings may help us to provide a prompt for feedback analysis focusing on clarification of the problem case based on individual analysis of the problem-solving partners' arguments (e.g., "okay_wait_how"; see Weinberger & Fischer, 2006). In contrast, the instructional task saw more use of cognitive processes and a "trial and error" approach, which led to disengagement, indicating that students who perceived below median peer support tended to express task understanding, motivation, planning, reflection, and co-regulation less frequently. Mapping these to the logged actions can help us refine methods for designing and developing collaborative scaffolds at appropriate junctures in model-building processes. For example, the student can support his claim to convince his partner to do a specific action with the supporting data (i.e., the student says "look at it" combined with OPEN DATA action).

7. Limitations and Future Work

This study has some limitations. First, the sample size was small, which limits our claims of generalizability. However, the small sample size also allowed us to perform a detailed multimodal learning analytics in a widely used, authentic (non-laboratory) OELE environment across different CPS tasks. This may not have been possible with a larger number of participants. In future work, we plan to extend this approach to examining data from a larger number of students. Second, our study also addresses how to monitor CPS processes over time, both within a single CPS session and across multiple sessions. Do teams improve regulation of CPS as the session progresses, and is the degree of improvement correlated to the outcomes? However, our results are correlational, not causal. An experimental design may help to make stronger claims. Another interesting study would involve comparing student behaviours, talk, and performance when they work individually (when using the think-aloud protocol) versus when they work in groups. Third, our NLP computational method is not intended to be a fully automated classification approach of regulation of CPS, nor is it meant to function as a solution to the problem of different NLP methods. Rather, our approach is a powerful aid to map the POS n-grams to our coding scheme (i.e., regulation of CPS coding system) and to the meaningful interpretation of computationally identified segments of the transcript. In future studies, we plan to automate the coding of verbal indicators. Researchers have begun investigating best practices for automated assessment using the PISA and ICAP frameworks (e.g., Hao, Chen, Flor, Liu, & von Davier, 2017; von Davier et al., 2017). Such automated coding could also be used as the basis to provide timely feedback to support students and teachers (Awwal, Scoular, & Alom, 2017). Future analysis could additionally be extended beyond analysis of verbal and log data by coding group-level regulation processes with respect to attunement, including back-channelling or nonverbal reactions (e.g., laughing, leaning in, eye contact; Grover et al., 2016; Isohätälä et al., 2017).

Overall, in this paper, by combining learning analytics and machine learning techniques, such as process mining and NLP, we have successfully performed a finer-grained analysis of student regulation processes during collaborative problem solving, and have shown how they impact their co-construction of knowledge, their model development process, and their learning. The CORTDRA diagrams provided us with visualizations of how student cognitive and regulation processes evolve over time. Clearly, student behaviours and regulation become more productive as they progress in a unit, and across units. Interestingly, the complexity of the tasks also influences student regulation and learning behaviours. These results provide actionable implications for designing collaborative activities involving coding in computer science and STEM classrooms.

Declaration of Conflicting Interest

The authors declared no potential conflicts of interest with respect to the research, authorship, and/or publication of this article.

Funding

This project was supported under National Science Foundation Award DRL-1640199.

Acknowledgments

We thank Kevin McElhaney, Satabdi Basu, Luke Conlin, Kristen Pilner Blair, Doris Chin, Rachel Wolf, Naveed Mohammed, Ramkumar Rajendran, Ahmed Elkhamary, and all our C2STEM project contributors for their assistance on this project.

References

Amon, M. J., Vrzakova, H., & D'Mello, S. K. (2019). Beyond dyadic coordination: Multimodal behavioral irregularity in groups predicts facets of collaborative problem solving. *Cognitive Science*, *43*(10), e12787. https://doi.org/10.1111/cogs.12787

Awwal, N., Scoular, C., & Alom, M. (2017). An automated system for evaluating 21st century skills using game-based assessments. *Proceedings of the 9th International Conference on Education and New Learning Technologies* (EDULEARN17), 3–5 July 2017, Barcelona, Spain (pp. 1593–1598). Valencia, Spain: International Academy of Technology, Education and Development (IATED).

Bakhtiar, A., & Hadwin, A. (2020). Dynamic interplay between modes of regulation during motivationally challenging episodes in collaboration. *Frontline Learning Research*, *8*(2), 1–34. https://doi.org/10.14786/flr.v8i2.561

Bakliwal, A., Arora, P., Patil, A., & Varma, V. (2011). Towards enhanced opinion classification using NLP techniques. *Proceedings of the Workshop on Sentiment Analysis Where AI Meets Psychology* (SAAIP 2011), November 2011, Chiang Mai, Thailand (pp. 101–107). Asian Federation of Natural Language Processing. https://www.aclweb.org/anthology/W11-3715

Bangerter, A. (2004). Using pointing and describing to achieve joint focus of attention in dialogue. *Psychological Science*, *15*(6), 415–419. https://doi.org/10.1111/j.0956-7976.2004.00694.x

Barron, B. (2003). When smart groups fail. *The Journal of the Learning Sciences*, *12*(3), 307–359.

Basu, S., Biswas, G., & Kinnebrew, J. S. (2016). Using multiple representations to simultaneously learn computational thinking and middle school science. *Proceedings of the 30th Conference on Artificial Intelligence* (AAAI-16), 12–17 February 2016, Phoenix, AZ, USA (pp. 3705–3711). Palo Alto, CA: AAAI Press.

Blackman, N. J.-M., & Koval, J. J. (2000). Interval estimation for Cohen's kappa as a measure of agreement. *Statistics in Medicine*, *19*(5), 723–741.

Broll, B., Lédeczi, A., Volgyesi, P., Sallai, J., Maroti, M., Carrillo, A., Weeden-Wright, S. L., Vanags, C., Swartz, J. D., & Lu, M. (2017). A visual programming environment for learning distributed programming. *Proceedings of the 48th ACM Technical Symposium on Computer Science Education* (SIGCSE '17), 8–11 March 2017, Seattle, WA, USA (pp. 81–86). New York: ACM.

Chiu, M. M., & Khoo, L. (2003). Rudeness and status effects during group problem solving: Do they bias evaluations and reduce the likelihood of correct solutions? *Journal of Educational Psychology*, *95*(3), 506.

Clark, H. H., & Brennan, S. E. (1991). Grounding in communication. In *Perspectives on socially shared cognition* (pp. 127–149). American Psychological Association. https://doi.org/10.1037/10096-006

De Backer, L., Van Keer, H., & Valcke, M. (2015). Exploring evolutions in reciprocal peer tutoring groups' socially shared metacognitive regulation and identifying its metacognitive correlates. *Learning and Instruction*, *38*, 63–78.

De Weerdt, J., De Backer, M., Vanthienen, J., & Baesens, B. (2012). A multi-dimensional quality assessment of state-of-the-art process discovery algorithms using real-life event logs. *Information Systems*, *37*(7), 654–676.

Dowell, N., Lin, Y., Godfrey, A., & Brooks, C. (2020). Exploring the relationship between emergent sociocognitive roles, collaborative problem-solving skills and outcomes: A group communication analysis. *Journal of Learning Analytics*, *7*(1), 38–57. https://doi.org/10.18608/jla.2020.71.4

Dowell, N. M., Nixon, T. M., & Graesser, A. C. (2019). Group communication analysis: A computational linguistics approach for detecting sociocognitive roles in multiparty interactions. *Behavior Research Methods*, *51*(3), 1007–1041.

Emara, M., Tscholl, M., Dong, Y., & Biswas, G. (2017). Analyzing students' collaborative regulation behaviors in a classroom-integrated open ended learning environment. In B. K. Smith, M. Borge, E. Mercier, & K. Y. Lim (Eds.), *Making a Difference: Prioritizing Equity and Access in CSCL: Proceedings of the 12th International Conference on Computer Supported Collaborative Learning* (CSCL 2017) 18–22 June 2017, Philadelphia, PA, USA (pp. 319–326). International Society of the Learning Sciences.

Emara, M., Grover, S., Hutchins, N. M., Biswas, G., & Snyder, C. (2020). Examining students' debugging and regulation processes during collaborative computational modeling in science. *14th International Conference of the Learning Sciences* (ICLS '20), 19–23 June 2020, Nashville, TN, USA. International Society of the Learning Sciences.

Ferreira, R., Kovanović, V., Gašević, D., & Rolim, V. (2018). Towards combined network and text analytics of student discourse in online discussions. In C. Penstein Rosé et al. (Eds.), *Artificial intelligence in education* (Proceedings of AIED 2018). Lecture Notes in Computer Science, vol. 10947. Springer, Cham. https://doi.org/10.1007/978-3-319-93843-1_9

Fischer, C., Pardos, Z. A., Baker, R. S., Williams, J. J., Smyth, P., Yu, R., Slater, S., Baker, R., & Warschauer, M. (2020). Mining big data in education: Affordances and challenges. *Review of Research in Education*, *44*(1), 130–160. https://doi.org/10.3102/0091732X20903304

Gašević, D., Jovanović, J., Pardo, A., & Dawson, S. (2017). Detecting learning strategies with analytics: Links with self-reported measures and academic performance. *Journal of Learning Analytics*, *4*(2), 113–128. https://doi.org/10.18608/jla.2017.42.10

Gobert, J., Slotta, J., Clarke, J., Dede, C., Gijlers, H., Saab, N., Van Joolingen, W., De Jong, T., & Koedinger, K. (2007). Fostering peer collaboration with technology. In C. Chinn, G. Erkens, & S. Puntambekar (Eds.), *Proceedings of the 7th International Conference on Computer-Supported Collaborative Learning* (CSCL 2007), 16–21 July 2007, New Brunswick, NJ, USA (pp. 23–27). International Society of the Learning Sciences.

Greene, J. A., & Azevedo, R. (2009). A macro-level analysis of SRL processes and their relations to the acquisition of a sophisticated mental model of a complex system. *Contemporary Educational Psychology*, *34*(1), 18–29.

Grover, S., Bienkowski, M., Tamrakar, A., Siddiquie, B., Salter, D., & Divakaran, A. (2016, April). Multimodal analytics to study collaborative problem solving in pair programming. *Proceedings of the 6th International Conference on Learning Analytics and Knowledge* (LAK '16), 25–29 April 2016, Edinburgh, UK (pp. 516–517). New York: ACM.

Grover, S., Basu, S., Bienkowski, M., Eagle, M., Diana, N., & Stamper, J. (2017). A framework for using hypothesis-driven approaches to support data-driven learning analytics in measuring computational thinking in block-based programming environments. *ACM Transactions on Computing Education (TOCE)*, *17*(3), 1–25.

Grover, S., Hutchins, N., Biswas, G., Snyder, C., & Emara, M. (2019). Examining synergistic learning of physics and computational thinking through collaborative problem solving in computational modeling. *Proceedings of the American Educational Research Association Annual Conference* (AERA 2019), 5–9 April 2019, Toronto, ON, Canada. https://www.researchgate.net/publication/332427176_Examining_Synergistic_Learning_of_Physics_and_Computational_Thinking_through_Collaborative_Problem_Solving_in_Computational_Modeling/stats

Grover, S., & Pea, R. (2013). Computational thinking in K–12: A review of the state of the field. *Educational Researcher*, *42*(1), 38–43.

Günther, C. W., & Van Der Aalst, W. M. (2007). Fuzzy mining: Adaptive process simplification based on multi-perspective metrics. In G. Alonso, P. Dadam, & M. Rosemann (Eds.), *Proceedings of the 5th International Conference on Business Process Management* (BPM 2007), 24–28 September 2007, Brisbane, Australia (pp. 328–343). Springer.

Hadwin, A., Järvelä, S., & Miller, M. (2018). Self-regulation, co-regulation, and shared regulation in collaborative learning environments. In D. H. Schunk & J. A. Greene (Eds.), Educational psychology handbook series. *Handbook of self-regulation of learning and performance* (pp. 83–106). Routledge/Taylor & Francis Group.

Hambrusch, S., Hoffmann, C., & Korb, J. T. (2009). A multidisciplinary approach towards computational thinking for science majors. *Proceedings of the 40th ACM Technical Symposium on Computer Science Education* (SIGCSE 09), 4–7 March 2009, Chattanooga, TN, USA (pp. 183–187). New York: ACM. https://doi.org/10.1145/1508865.1508931

Hao, J., Chen, L., Flor, M., Liu, L., & von Davier, A. A. (2017). CPS-Rater: Automated sequential annotation for conversations in collaborative problem-solving activities. *ETS Research Report Series, 2017*(1). https://doi.org/1-9. 10.1002/ets2.12184

Henderson, P. B., Cortina, T. J., & Wing, J. M. (2007). Computational thinking. *Proceedings of the 38th ACM Technical Symposium on Computer Science Education* (SIGCSE 07), 7–10 March 2007, Covington, KY, USA (pp. 195–196). New York: ACM.

Hmelo-Silver, C. E., Chernobilsky, E., & Jordan, R. (2008). Understanding collaborative learning processes in new learning environments. *Instructional Science*, *36*(5–6), 409–430.

Hutchins, N., Biswas, G., Grover, S., Basu, S., & Snyder, C. (2019). A systematic approach for analyzing students' computational modeling processes in C2STEM. In S. Isotani, E. Millán, A. Ogan, P. Hastings, B. McLaren, & R. Luckin (Eds.), *Proceedings of the 20th International Conference on Artificial Intelligence in Education* (AIED 2019), 25–29 June 2019, Chicago, IL, USA (pp. 116–121). Springer. https://doi.org/10.1007/978-3-030-23207-8_22

Hutchins, N. M., Biswas, G., Zhang, N., Snyder, C., Lédeczi, Á., & Maróti, M. (2020a). Domain-specific modeling languages in computer-based learning environments: A systematic approach to support science learning through computational modeling. *International Journal of Artificial Intelligence in Education*, *30*, 537–580. https://doi.org/10.1007/s40593-020-00209-z

Hutchins, N. M., Biswas, G., Maróti, M., Lédeczi, Á., Grover, S., Wolf, R., Blair, K. P., Chin, D., Conlin, L., Basu, S., & McElhaney, K. (2020b). C2STEM: A system for synergistic learning of physics and computational thinking. *Journal of Science Education and Technology, 29*(1), 83–100. https://doi.org/10.1007/s10956-019-09804-9

Iiskala, T., Vauras, M., Lehtinen, E., & Salonen, P. (2011). Socially shared metacognition of dyads of pupils in collaborative mathematical problem-solving processes. *Learning and Instruction*, *21*(3), 379–393. https://doi.org/10.1016/j.learninstruc.2010.05.002

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., 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.

Jona, K., Wilensky, U., Trouille, L., Horn, M. S., Orton, K., Weintrop, D., & Beheshti, E. (2014). Embedding computational thinking in science, technology, engineering, and math (CT-STEM). *Future Directions in Computer Science Education Summit Meeting*, Orlando, FL.

Kapur, M. (2016). Examining productive failure, productive success, unproductive failure, and unproductive success in learning. *Educational Psychologist*, *51*(2), 289–299. https://doi.org/10.1080/00461520.2016.1155457

Kinnebrew, J., Segedy, J. R., & Biswas, G. (2017). Integrating model-driven and data-driven techniques for analyzing learning behaviors in open-ended learning environments. *IEEE Transactions on Learning Technologies*, *10*(2), 140–153. https://doi.org/10.1109/TLT.2015.2513387

Klahr, D., & Carver, S. M. (1988). Cognitive objectives in a LOGO debugging curriculum: Instruction, learning, and transfer. *Cognitive Psychology, 20*(3), 362–404.

Kneser, C., Pilkington, R., & Treasure-Jones, T. (2001). The tutor's role: An investigation of the power of exchange structure analysis to identify different roles in CMC seminars. *International Journal of Artificial Intelligence in Education*, *12*(63–84).

Kurniati, A. P., Kusuma, G., & Wisudawan, G. (2016). Implementing heuristic miner for different types of event logs. *International Journal of Applied Engineering Research, 11*, 5523–5529. https://www.researchgate.net/publication/301998930_Implementing_Heuristic_Miner_for_Different_Types_of_Event_Logs

Landau, R. (2006). Computational physics: A better model for physics education? *Computing in Science & Engineering*, *8*(5), 22–30.

Lin, K.-Y., Yu, K.-C., Hsiao, H.-S., Chu, Y.-H., Chang, Y.-S., & Chien, Y.-H. (2015). Design of an assessment system for collaborative problem solving in STEM education. *Journal of Computers in Education*, *2*(3), 301–322. https://doi.org/10.1007/s40692-015-0038-x

Liu, Z., Zhi, R., Hicks, A., & Barnes, T. (2017). Understanding problem solving behavior of 6–8 graders in a debugging game. *Computer Science Education*, *27*(1), 1–29. https://doi.org/10.1080/08993408.2017.1308651

Lobczowski, N. G., Allen, E. M., Firetto, C. M., Greene, J. A., & Murphy, P. K. (2020). An exploration of social regulation of learning during scientific argumentation discourse. *Contemporary Educational Psychology*, 101925. https://doi.org/10.1016/j.cedpsych.2020.101925

Loksa, D., & Ko, A. J. (2016). The role of self-regulation in programming problem solving process and success. *Proceedings of the 12th Annual Conference on International Computing Education Research* (ICER 2016), 8–12 September 2016, Melbourne, Australia (pp. 83–91). New York: ACM. https://doi.org/10.1145/2960310.2960334

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.

Malmberg, J., Järvelä, S., & Kirschner, P. A. (2014). Elementary school students' strategic learning: Does task-type matter? *Metacognition and Learning*, *9*(2), 113–136.

McNamara, D. S. (2011). Computational methods to extract meaning from text and advance theories of human cognition. *Topics in Cognitive Science*, *3*(1), 3–17. https://doi.org/10.1111/j.1756-8765.2010.01117.x

Mohammed, M., & Omar, N. (2020). Question classification based on Bloom's taxonomy cognitive domain using modified TF-IDF and word2vec. *PloS One*, *15*(3), e0230442. https://doi.org/10.1371/journal.pone.0230442

Molenaar, I., Chiu, M. M., Sleegers, P., & van Boxtel, C. (2011). Scaffolding of small groups' metacognitive activities with an avatar. *International Journal of Computer-Supported Collaborative Learning*, *6*(4), 601–624. https://doi.org/10.1007/s11412-011-9130-z

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. https://doi.org/10.1007/s11409-013-9105-8

Noroozi, O., Alikhani, I., Järvelä, S., Kirschner, P. A., Juuso, I., & Seppänen, T. (2019). Multimodal data to design visual learning analytics for understanding regulation of learning. *Computers in Human Behavior*, *100*, 298–304.

Oshima, J., Tsunakawa, T., & Oshima, R. (2020). An assessment of idea emergence in subject-matter collaborative learning. *Frontiers in Education*, *5*, 21.

Paans, C., Onan, E., Molenaar, I., Verhoeven, L., & Segers, E. (2019). How social challenges affect children's regulation and assignment quality in hypermedia: A process mining study. *Metacognition and Learning*, *14*(2), 189–213.

Papert, S., & Harel, I. (1991). Situating constructionism. *Constructionism, 36*(2), 1–11.

Peters-Burton, E. E., Cleary, T. J., & Kitsantas, A. (2018). Computational thinking in the context of science and engineering practices: A self-regulated learning approach. *Digital Technologies: Sustainable Innovations for Improving Teaching and Learning* (pp. 223–240). Springer.

Pino-Pasternak, D., Whitebread, D., & Neale, D. (2018). The role of regulatory, social, and dialogic dynamics on young children's productive collaboration in group problem solving. *New Directions for Child and Adolescent Development*, *2018*(162), 41–66.

Praharaj, S., Scheffel, M., Drachsler, H., & Specht, M. (2018). Multimodal analytics for real-time feedback in co-located collaboration. *Proceedings of the 13th European Conference on Technology Enhanced Learning* (EC-TEL 2018), 3–5 September 2018, Leeds, UK (pp. 187–201). Lecture Notes in Computer Science, vol. 11082, Springer.

Redish, E. F., & Wilson, J. M. (1993). Student programming in the introductory physics course: M.U.P.P.E.T. *American Journal of Physics*, *61*(3), 222–232. https://doi.org/10.1119/1.17295

Repenning, A., Webb, D., & Ioannidou, A. (2010). Scalable game design and the development of a checklist for getting computational thinking into public schools. *Proceedings of the 41st ACM Technical Symposium on Computer Science Education* (SIGCSE 10), 10–13 March 2010, Milwaukee, WI, USA (pp. 265–269). New York: ACM. https://doi.org/10.1145/1734263.1734357

Roschelle, J., & Teasley, S. D. (1995). The construction of shared knowledge in collaborative problem solving. In J. L. Schnase & E. L. Cunnius (Eds.), *Proceedings of the 1st International Conference on Computer-Support for Collaborative Learning* (CSCL '95), 17–20 October 1995, Bloomington, IN, USA (pp. 69–97). International Society of the Learning Sciences. Mahwah, NJ: Lawrence Erlbaum Associates.

Rosé, C., Wang, Y.-C., Cui, Y., Arguello, J., Stegmann, K., Weinberger, A., & Fischer, F. (2008). Analyzing collaborative learning processes automatically: Exploiting the advances of computational linguistics in computer-supported collaborative learning. *International Journal of Computer-Supported Collaborative Learning*, *3*(3), 237–271.

Schneider, B., Sharma, K., Cuendet, S., Zufferey, G., Dillenbourg, P., & Pea, R. (2018). Leveraging mobile eye-trackers to capture joint visual attention in co-located collaborative learning groups. *International Journal of Computer-Supported Collaborative Learning*, *13*(3), 241–261. https://doi.org/10.1007/s11412-018-9281-2

Schoor, C., & Bannert, M. (2012). Exploring regulatory processes during a computer-supported collaborative learning task using process mining. *Computers in Human Behavior*, *28*(4), 1321–1331.

Segedy, J. R., Kinnebrew, J. S., & Biswas, G. (2015). Using coherence analysis to characterize self-regulated learning behaviours in open-ended learning environments. *Journal of Learning Analytics*, *2*(1), 13–48. https://doi.org/10.18608/jla.2015.21.3

Sengupta, P., Kinnebrew, J. S., Basu, S., Biswas, G., & Clark, D. (2013). Integrating computational thinking with K–12 science education using agent-based computation: A theoretical framework. *Education and Information Technologies*, *18*(2), 351–380.

Siadaty, M., Gašević, D., & Hatala, M. (2016). Associations between technological scaffolding and micro-level processes of self-regulated learning: A workplace study. *Computers in Human Behavior*, *55*, 1007–1019.

Snyder, C., Hutchins, N., Biswas, G., Emara, M., Grover, S., & Conlin, L. (2019). Analyzing students' synergistic learning processes in physics and CT by collaborative discourse analysis. In K. Lund, G. Niccolai, E. Lavoué, C. Hmelo-Silver, G. Gweon, & M. Baker (Eds.), *A Wide Lens: Combining Embodied, Enactive, Extended, and Embedded Learning in Collaborative Settings. Proceedings of the 13th International Conference on Computer Supported Collaborative Learning* (CSCL 2019), 17–21 June 2019, Lyon, France (pp. 360–367). International Society of the Learning Sciences. https://doi.dx.org/10.22318/cscl2019.360

Sobocinski, M., Järvelä, S., Malmberg, J., Dindar, M., Isosalo, A., & Noponen, K. (2020). How does monitoring set the stage for adaptive regulation or maladaptive behavior in collaborative learning? *Metacognition and Learning, 15,* 99–127. https://doi.org/10.1007/s11409-020-09224-w

Sobocinski, M., Malmberg, J., & Järvelä, S. (2017). Exploring temporal sequences of regulatory phases and associated interactions in low-and high-challenge collaborative learning sessions. *Metacognition and Learning*, *12*(2), 275–294.

Soderstrom, N. C., & Bjork, R. A. (2015). Learning versus performance: An integrative review. *Perspectives on Psychological Science*, *10*(2), 176–199. https://doi.org/10.1177/1745691615569000

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. https://doi.org/10.18608/jla.2015.21.5

Stewart, A. E., Vrzakova, H., Sun, C., Yonehiro, J., Stone, C. A., Duran, N. D., . . ., & D'Mello, S. K. (2019). I say, you say, we say: Using spoken language to model socio-cognitive processes during computer-supported collaborative problem solving. *Proceedings of the ACM on Human–Computer Interaction* (Volume 3, Issue CSCW, Article No. 194, pp. 1–19), November 2019. New York: ACM. https://doi.org/10.1145/3359296

Sullivan, F. R., & Keith, P. K. (2019). Exploring the potential of natural language processing to support microgenetic analysis of collaborative learning discussions. *British Journal of Educational Technology*, *50*(6), 3047–3063.

Sun, C., Shute, V. J., Stewart, A., Yonehiro, J., Duran, N., & D'Mello, S. (2020). Towards a generalized competency model of collaborative problem solving. *Computers, & Education*, *143*, 103672. https://doi.org/10.1016/j.compedu.2019.103672

Toulmin, S. E. (2001). *Return to reason*. Cambridge, MA: Harvard University Press.

Van den Bossche, P., Gijselaers, W., Segers, M., Woltjer, G., & Kirschner, P. (2011). Team learning: Building shared mental models. *Instructional Science*, *39*(3), 283–301. https://doi.org/10.1007/s11251-010-9128-3

Van der Aalst, W. M. (2011). Process discovery: An introduction. In *Process mining: Discovery, conformance and enhancement of business processes* (pp. 125–156). Springer. https://doi.org/10.1007/978-3-642-19345-3_5

Verheij, B. (2005). Evaluating arguments based on Toulmin's scheme. *Argumentation*, *19*(3), 347-371. https://doi.org/10.1007/s10503-005-4421-z

von Davier, A. A., Hao, J., Liu, L., & Kyllonen, P. (2017). Interdisciplinary research agenda in support of assessment of collaborative problem solving: Lessons learned from developing a collaborative science assessment prototype. *Computers in Human Behavior*, *76*, 631–640. https://doi.org/10.1016/j.chb.2017.04.059

Weijters, A. J. M. M., van Der Aalst, W. M., & De Medeiros, A. A. (2006). Process mining with the heuristics miner-algorithm. *Technische Universiteit Eindhoven, Tech. Rep. WP*, *166*, 1–34.

Weinberger, A., & Fischer, F. (2006). A framework to analyze argumentative knowledge construction in computer-supported collaborative learning. *Computers and Education, 46*(1), 71–95. https://doi.org/10.1016/j.compedu.2005.04.003

Weintrop, D., Beheshti, E., Horn, M., Orton, K., Jona, K., Trouille, L., & Wilensky, U. (2016). Defining computational thinking for mathematics and science classrooms. *Journal of Science Education and Technology*, *25*(1), 127–147. https://doi.org/10.1007/s10956-015-9581-5

Werner, L., McDowell, C., & Denner, J. (2013). A first step in learning analytics: Pre-processing low-level Alice logging data of middle school students. *Journal of Educational Data Mining, 5*(2), 11–37. https://doi.org/10.5281/zenodo.3554631

Williams, L., Wiebe, E., Yang, K., Ferzli, M., & Miller, C. (2002). In support of pair programming in the introductory computer science course. *Computer Science Education*, *12*(3), 197–212.

Winne, P. H., & Hadwin, A. F. (1998). Studying as self-regulated learning. *Metacognition in Educational Theory and Practice*, *93*, 27–30.

Winters, F. I., & Alexander, P. A. (2011). Peer collaboration: The relation of regulatory behaviors to learning with hypermedia. *Instructional Science*, *39*(4), 407–427. https://doi.org/10.1007/s11251-010-9134-5

Wise, A. F., Azevedo, R., Stegmann, K., Malmberg, J., Rosé, C. P., Mudrick, N., Taub, M., Martin, S. A., Farnsworth, J., & Mu, J. (2015). *CSCL and learning analytics: Opportunities to support social interaction, self-regulation and socially shared regulation*. International Society of the Learning Sciences.

Worsley, M., & Blikstein, P. (2011). What's an expert? Using learning analytics to identify emergent markers of expertise through automated speech, sentiment and sketch analysis. In M. Pechenizkiy et al. (Eds.), *Proceedings of the 4th Annual Conference on Educational Data Mining* (EDM2011), 6–8 July 2011, Eindhoven, Netherlands (pp. 235–240). International Educational Data Mining Society.

Xie, K., Di Tosto, G., Lu, L., & Cho, Y. S. (2018). Detecting leadership in peer-moderated online collaborative learning through text mining and social network analysis. *The Internet and Higher Education*, *38*, 9–17.

Yett, B., Hutchins, N., Stein, G., Zare, H., Snyder, C., Biswas, G., Metelko, M., & Lédeczi, Á. (2020). A hands-on cybersecurity curriculum using a robotics platform. *Proceedings of the 51st ACM Technical Symposium on Computer Science Education* (SIGCSE '20), 11–14 March 2020, Portland, OR, USA (pp. 1040–1046). New York: ACM.