Measuring Students' Self-Regulatory Phases in LMS with Behavior and Real-Time Self Report

Fatemeh Salehian Kia
Simon Fraser University
University of Michigan
fsalehia@sfu.ca

Marek Hatala
Simon Fraser University
mhatala@sfu.ca

Ryan S. Baker
University of Pennsylvania
ryanshaunbaker@gmail.com

Stephanie D. Teasley
University of Michigan
steasley@umich.edu

ABSTRACT

Research has emphasized that self-regulated learning (SRL) is critically important for learning. However, students have different capabilities of regulating their learning processes and individual needs. To help students improve their SRL capabilities, we need to identify students' current behaviors. Specifically, we applied instructional design to create visible and meaningful markers of student learning at different points in time in LMS logs. We adopted knowledge engineering to develop a framework of proximal indicators representing SRL phases and evaluated them in a quasi-experiment in two different learning activities. A comparison of two sources of collected students' SRL data, self-reported and trace data, revealed a relatively high agreement between our classifications (weighted kappa, κ = .74 and κ = .68). However, our indicators did not always discriminate adjacent SRL phases, particularly for enactment and adapting phases, compared with students' real-time self-reported behaviors. Our behavioral indicators also were comparably successful at classifying SRL phases for different self-regulatory engagement levels. This study demonstrated how the triangulation of various sources of students' self-regulatory data could help to unravel the complex nature of metacognitive processes.

CCS CONCEPTS

• **Information systems → Task models**; • **Human-centered computing → Empirical studies in HCI**; • **Applied computing → Learning management systems**.

KEYWORDS

self-regulated learning, pattern recognition, knowledge-engineered trace measures, self-reported measures

LAK21, April 12–16, 2021, Irvine, CA, USA

ACM ISBN 978-1-4503-8935-8/21/04…$15.00
https://doi.org/10.1145/3448139.3448164

ACM Reference Format:
Fatemeh Salehian Kia, Marek Hatala, Ryan S. Baker, and Stephanie D. Teasley. 2021. Measuring Students' Self-Regulatory Phases in LMS with Behavior and Real-Time Self Report. In *LAK21: 11th International Learning Analytics and Knowledge Conference (LAK21), April 12–16, 2021, Irvine, CA, USA.* ACM, New York, NY, USA, 11 pages. https://doi.org/10.1145/3448139.3448164

1 INTRODUCTION

Self-regulated learning (SRL) is a critical component of student learning as research has emphasized that it is critically related to academic performance [16] and regarded as a necessary skill for life-long learning [18]. However, students differ in their self-regulatory learning skills; they have different capabilities, know and adopt different strategies and tactics to study. Research showed that students can learn how to become self-regulated learners [27]. Supporting students with individual self-regulatory needs requires identifying students' current metacognitive abilities.

The increasing use of technology in education and the resulting ability to keep track of learning activities offer the opportunity to understand students' behaviors at an unprecedented level of analysis and scale. A few studies have adopted methods such as sequence analysis [14, 19], or process mining [5, 6] to detect self-regulatory behaviors from students' trace data in online learning environments. Despite recent advances in detecting students' self-regulatory behaviors, the research has focused more on advancing analytical methods, rather than examining the extent to which students' observed online activities, known as proximal indicators, contribute to a valid inference of self-regulated learning phases. In addition, the methods applied to detect behavioral indicators need to be tested for the extent and conditions to detect self-regulatory phases. A recent study [24] compared the results of three data analyses, i.e. sequence, process, and network analytics, to detect learning tactics and strategies in MOOCs. Their findings indicated differences in detecting learning tactics from the same trace data between the three different analytical methods.

In this paper, we developed knowledge-engineered indicators of students' self-regulatory phases in the LMS to examine the extent to which the observed online activities represent valid inference of SRL phases. We compared behavioral indicators with concurrent students' self-reported behaviors while performing a learning task to test the extent and conditions these indicators and self-reported measures converge. Since retrospective questionnaires allow participants to rely more on actual behaviors, rather than assessing metacognitive knowledge prospectively [11], we used retrospective self-reported measure of SRL to categorize students' self-regulatory engagement level. We have examined the differences between behavioral and concurrent self-reported measures among groups with different SRL levels. The triangulation of multiple data sources allowed us to create a more accurate picture of different students' self-regulated learning behaviors.

1.1 Background

Traditionally, the self-regulatory aspects of students' metacognition have been measured using questionnaires. However, research on SRL measures has shown that learners can be inaccurate in calibrating their learning behaviors [32, 44]. There is contention among some researchers [15] that students are not accurate reporters of their behaviors and therefore we should question the validity of self-reported measures. On the other hand, other researchers e.g. [20] emphasize the importance of understanding students' conception of themselves. Since disagreements exist regarding SRL measurements, particularly whether self-reports represent a valid and reliable approach to measuring these processes, researchers have advocated the use of behavioral data [44]. Therefore, trace data has been used to detect students' learning behaviors in virtual learning environments. This approach has provided promising insight into learning processes as an alternative to traditional approaches for measuring self-regulated learning.

A study by [43] is among early research on collecting students' trace data to measure self-regulatory phases. They used log data of software called *PrepMate* designed for authoring instructional presentations to capture students' interactions with reading materials. For instance, scrolling through a paragraph before starting any other actions was used as a proximal indicator of students' *planning* phase. A recent study by [8] used students' access to resources related to course organizations such as exercise deadlines or index page with learning objectives in the LMS as proximal indicators of students' *planning* phase. Although these studies detected students' self-regulatory phases in different contexts, the proximal indicators of *planning* phase were quite different in terms of their operationalization. The former used a learning tactic of skimming as a representation of *planning*, whereas the latter used indicators identifying students' *planning* strategy, i.e. organization and goal setting. Thus, there is a need for a consistent approach to developing proximal indicators that operationalize self-regulatory phases in online learning informed by theories of self-regulated learning.

Furthermore, SRL is a context-dependent process [31]. Capturing the context of students' actions in virtual learning environments is more feasible when learning activities occur in well-defined tasks. This is not the case for open-ended asynchronous learning environments such as LMS or MOOCs. Most prior research on identifying self-regulatory phases was conducted within well-defined tasks in advanced learning technologies (ALTs), such as intelligent tutoring systems, simulations, hypermedia, and serious games. For instance, Hadwin et al. [17] conducted an exploratory case study to examine in depth the students' learning activities across studying episodes in a learning environment, *gStudy*, designed specifically for this purpose. However, examining a studying episode in open-ended environments, particularly LMS, is not the same because of the necessarily content-agnostic nature of LMS [22], which makes it more difficult to capture the context of students' actions. Similar observable activities in the LMS may represent different behaviors in different learning contexts. Recent research on detecting SRL behaviors in open-ended environments has shifted away from using the frequency of students' activities to observing students' sequences of actions to better understand the context of students' choice of actions. A study by [19] is an example of adopting the combination of exploratory sequence analysis and hierarchical clustering to detect patterns in students' behaviors that indicate learning strategies in the LMS. However, the sequences may be misrepresented as evidence of the same behavior and grouped in the same cluster without accounting for the context surrounding these sequences. Another study by [38] used their coherence analysis approach [39] to capture the surrounding context of a sequence of actions. For instance, they considered two actions as coherent if the second action was based on information generated by the first action.

Several interdisciplinary researchers have addressed the complex nature of temporally unfolding SRL processes by using multichannel trace data such as log files, eye-tracking, physiological sensors, and screen recordings of learners' interactions with machines [2]. For instance, a research study by [23] collected physiological data, video observations, and facial recognition data in the context of collaborative learning to explore how different sources of data can be used to detect self-regulatory components of students' interactions. However, due to the specific LMS affordances, we only can track a glimpse of learning processes that mostly occur outside the LMS. One of the common approaches often used in SRL measurement settings is concurrent think-aloud protocols [12]. A recent study by [40] integrated coded think-aloud and trace data and applied process mining to understand students' self-regulatory behaviors. A similar approach, such as collecting students' real-time self-reported behaviors in the LMS, can provide a detailed trace of the learning process. Additionally, there is evidence that converging self-reported and process data is key to understanding metacognitive self-regulatory processes [25].

This study contributes to the growing research of SRL modeling in several ways. First, we applied instructional design to create a template for a prevalent learning activity in the LMS not only to support students' self-regulatory process but also to create visible and meaningful markers of student learning at different points in time in LMS logs. Second, we developed indicators of self-regulated learning phases from students' trace data by applying a multi-method approach. Third, we developed and embedded a tool in the LMS to collect students' self-reported learning behaviors in real-time. Finally, to test our indicators' robustness in identifying self-regulatory phases in the LMS, our research design allowed us to empirically examine the extent to which proximal indicators converge with their corresponding student self-reported behaviors. Our study highlights the importance of course design in supporting learning analytics to uncover the context of students' learning behaviors. Moreover, our research design demonstrated how the triangulation of different sources of students' self-regulatory data can help to unravel the complex nature of metacognitive processes. Finally, our empirical findings showed how the indicators of self-regulation were relatively robust in various learning contexts for different students' self-regulatory engagement levels.

1.2 Research Questions

This study is guided by the following research questions:

RQ1. *How can we identify proximal indicators of SRL using log tracing techniques, given specific affordances of the LMS?*

RQ2. *What specific approach leads to collecting students' self-reported behaviors in real-time to minimize the concerns associated with students' reporting their metacognition?*

RQ3. *To what extent do observed indicators of SRL and students self-reported behaviors converge?*

2 METHODS

2.1 Learning Activity Design

Since an LMS is an open-ended learning environment, instructional design plays a vital role in getting students engaged in self-directed learning activities such as individual assignments. Gašević et al. [13] emphasized the impact of differences in instructional conditions, particularly those related to whether and how to use LMS features, for creating a generalized predictive model of students' success. Moreover, Rienties et al. [30] indicated that the design of learning activities influenced students' engagement in online learning. Therefore, in this study, we focused on a specific learning activity designed to engage students in a complex problem-solving task, which requires students to be self-regulated learners [36]. The students were provided with information about a computer programming task in the LMS Assignment module. They were required to integrate theory and practice and apply knowledge and skills to develop a viable solution to a given problem over two weeks. Although the assignment instructions can be given in different formats, such as a downloadable PDF document or a single HTML page in the LMS (Canvas), we designed a template incorporating *task specific sub-goals* (TSSG) [37]. Accordingly, the assignment instructions consisted of six sections, i.e. *Overview, General Guidelines, Detail Specification, Resources, Marking Scheme,* and *Submission Instructions* (see below). The instructions provided information for students explaining specific, measurable, action-oriented and realistic steps to complete the assignment, and allowed us to track their progress by when and which kind of information they viewed. Placing each kind of information into its separate page brings an advantage of utilizing the built-in logging mechanism of Canvas, as opening the assignment front page (the organizer) and each of the pages (sections) creates a separate LMS log entry. These log entries carry semantic meaning, which we used to develop indicators of self-regulatory phases. The assignment front page in Canvas, partly seen in Figure 3, provided the links to six content pages with the following information:

- *Overview* explains the learning objectives of the assignment.
- *General Guidelines* tells learners how to approach the problem by dividing it into sub-tasks, indicating what knowledge and techniques to use.
- *Detail Specification* describes technical requirements for the task, i.e. functional requirements for the solution.
- *Resources* provides the list of supplementary resources, e.g. links to the relevant lecture notes, technical documentation, and worked examples.
- *Marking Scheme* clarifies how the solution will be evaluated based on the accuracy and completeness of the learning product.
- *Submission Instructions* explains submission format requirements.

2.2 Developing Indicators of Self-Regulated Learning

To develop SRL indicators for the problem-solving task utilizing our design, in the Spring 2019 term we collected LMS interaction data from 92 students in an undergraduate computer programming course at a large residential research university in Canada. The log data included every click event in Canvas, including students' interactions with the six sections of the assignment (described in the previous section), downloading handout files, accessing worked examples, and discussion activities regarding the assignment. To develop the indicators, we have identified three different assignments in which we examined the sequences of 47 students (51%), until we reached saturation [35]. The students used an integrated development environment (IDE) to develop the computer program outside Canvas offline, which was not tracked. However, students' interactions with the assignment prompt pages, and other related modules in Canvas, generated sequences of events represented in the self-regulatory phases.

To develop indicators for our domain, we used a lens of the well-established SRL model by Winne and Hadwin [42] because of its information-processing perspective, since we detect students' behaviors from interactions with the information provided either as task-specific sub-goals or learning resources for them. Winne and Hadwin's model defines a loosely sequenced cycle of four phases to perform an academic task: task definition, planning and goal setting, enactment of tactics and strategies, and adapting. In each of these phases, learners find themselves in a set of processes involving interaction between the conditions, operations, products, evaluations, and standards (COPES) [41]. First, we distilled LMS logs to those clicks related to the assignment. Next, we applied a text replay tagging technique [3, 34]. A human expert carefully examined the logs recreating the sequence of clicks through the pages and evaluated the sequences for possible representations of SRL phases using the theoretical framework above. For each phase, we created a set of indicators, i.e. frequent sub-sequences, consisting of event clicks within a time frame representing self-regulatory phases. This process led to several findings related to our process itself, as described in the following subsections.

2.2.1 The importance of the context surrounding the sequence. Our qualitative analyses of LMS logs suggest that the context surrounding the same learning events' sequence plays an important role in discriminating sequences of students' actions as representations of self-regulatory phases. The example (Figure 1) shows the same sequence of LMS events, which can be the indicators of two different self-regulatory phases. The first occurrence of (View_Assignment_Objective, View_Assignment) can be an indicator of planning phase while the second occurrence of the same sequence may not represent the same self-regulatory phase because of the long elapsed time (3 days) between these two sequences, and the student downloaded a worked example and viewed a handout for another week between them. Thus, this sequence can be an indicator of *enactment* phase rather than *planning*.

student_id	timestamp	event	extra
subject_001	2019-02-02 17:27:35	VIEW_File	lecture_week_2
subject_001	2019-02-02 17:29:45	VIEW_ASSIGNMENT_OBJECTIVE	assignment_1
subject_001	2019-02-02 17:31:56	VIEW_ASSIGNMENT	assignment_1
subject_001	2019-02-02 17:45:21	DOWNLOAD_FILE	worked_example
subject_001	2019-02-03 19:13:22	VIEW_File	lecture_week_3
subject_001	2019-02-05 23:21:51	VIEW_ASSIGNMENT_OBJECTIVE	assignment_1
subject_001	2019-02-05 23:22:03	VIEW_ASSIGNMENT	assignment_1

Figure 1: The example of the same sequence representing different behaviors in trace data

2.2.2 The macro-level order of sequences matters. The temporality of self-regulated learning phases should determine the granularity level in which we create a sequence of actions for every student. That is, the micro-level order of events is not a deterministic factor in detecting self-regulatory phases if the subsequent events occur within a short time frame. The example in Figure 1 shows subsequent events (View_File, View_Assignment_Objective, View_Assignment, Download_File) that occurred within a few minutes. The order of these events does not matter in terms of identifying self-regulatory phases; rather, what occurred in this time frame can help to determine which self-regulatory phase these events would represent or provoke. As a result, we propose to segment click streams into *sessions* or clips. A session is the grain-size of a time frame at which a self-regulatory indicator can be detected. Hence, we divided the students' interactions with the assignment and related modules in the LMS into sessions. The sessions allowed us to differentiate the same sequence of events that occurred in different time frames, representing two different learning behaviors and accounting for the temporality of the underlying self-regulatory phases. We defined a 20-minute cutoff for each session, determined by the students' LMS use data, which was clustered in this time frame with a typical length of ~15 minutes. The events within a session demonstrated:

- access to information about task conditions was determined as indicators of a *"task definition"* phase,
- first time access to information concerning task operations and learning objectives represented a *"planning"* phase,
- access to information regarding task operations which was followed by an action outside the LMS, writing a computer program in the IDE was used as indicators of an *"enactment"* phase,
- access to information about assessment criteria or task standards that invoked making evaluations and changes of the computer program and adapting the product was identified as indicators of an *"adapting"* phase.

The list of indicators for each of self-regulatory phases is shown in Figure 2.

2.3 Procedure

In the Summer 2020 term, we conducted a quasi-experiment in the next offering of the same course used to develop the indicators, taught by the same instructor. A researcher from our team invited the students to participate in the study and explained the procedure. Participation in the study was voluntary; students were informed that their responses would not be visible or reported to their course instructor. The participants received a bonus of 2% towards their final course grade. The participants were to work on the assignment as other students, for two weeks, at any time they chose. Each time they looked at the assignment related pages, they were prompted at specific points during their learning episodes to respond to a single question asking what they were doing on the assignment. All participants' interactions with Canvas, as well as their self-reported responses, were timestamped and logged. At the end of the study, the participants were asked to fill out a survey about self-regulation aspects of their metacognition.

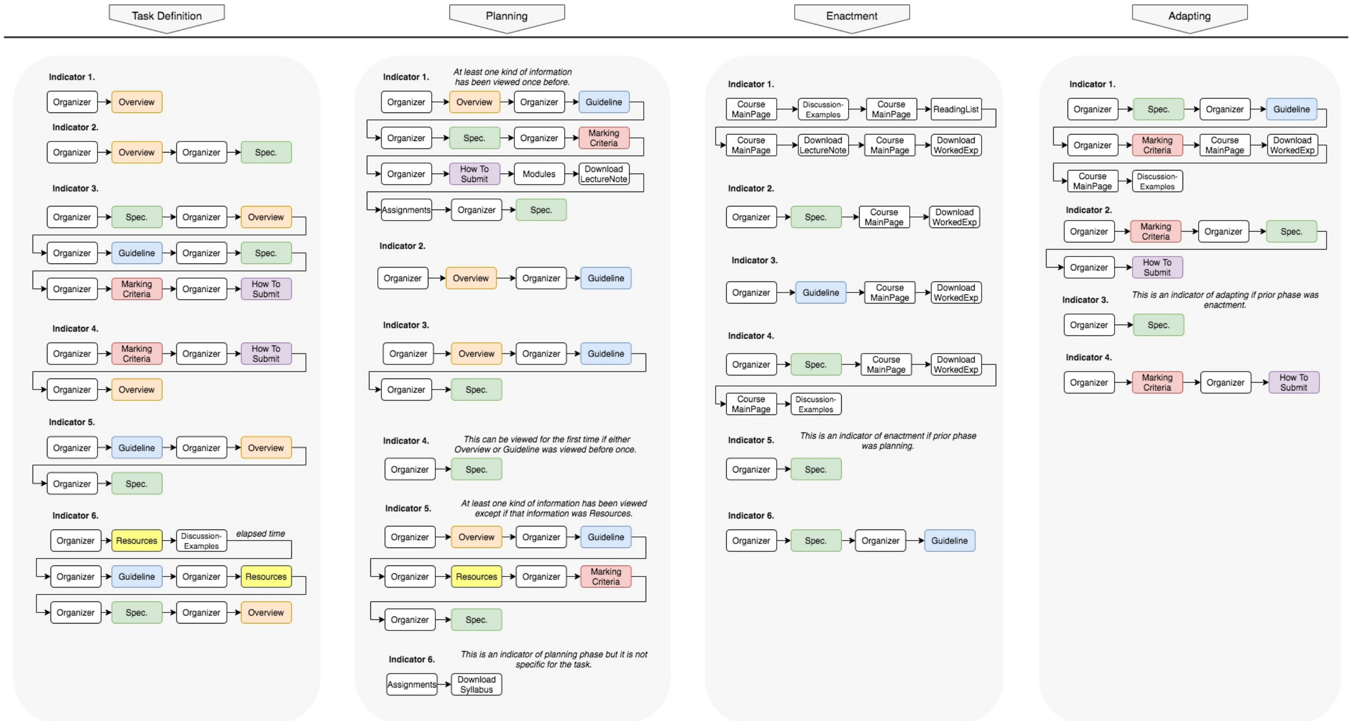
2.4 Participants

We recruited 38 out of 66 (58%) students to participate in our study (54% female, 44% male, and 2% undisclosed). Two consecutive assignments, Assignment 2 and 3, were chosen to collect participants' self-reported behaviors. The assignment tasks were different, but all assignments targeted to develop students' problem-solving skills. To minimize the learning effect, which could potentially influence students' clickstream trace, the students became familiar with the assignment structure in Assignment 1, which followed the same design as Assignments 2 and 3. Assignment 2 was available to students between weeks 3-5 and Assignment 3 between weeks 6-8 of a 13-week term. 21 out of 38 students self-reported their computing proficiency as meeting expectations for the course materials, 13 below expectations, and 4 exceeded expectations. After six students dropped the course between Assignments 2 and 3, we collected data for Assignment 3 from a total of 32 students.

2.5 Collecting Real-Time Self-Reported Behaviors

Self-reported questionnaires are a form of measure that is easy to collect, however, there are two main concerns regarding students' self-reported measures. First, self-regulated learning is considered to be a context-dependent process [41], and it may vary both across and within learning tasks and contexts. The second concern is the question of whether students can self-report their metacognitive behaviors, because of students' imperfect memory, inclination to provide socially desirable answers, or lack of awareness or control of their metacognition [32]. Due to these concerns, researchers have increasingly advocated the use of other forms of measures such as a multi-method approach by incorporating short micro-analytic questionnaires at various points during learning episodes [9].

To address the above concerns, we developed an embedded pop-up survey tool in the LMS (Canvas). The students were prompted to answer a question (see Figure 3). The prompt (orange box) appeared at the beginning of a student's interaction with any of the assignment pages; the question was asked again every 20 minutes after the student's last answer, as long as the student kept any of the assignment pages open in any of the browser tabs. This schedule was designed based on our findings in Section 2.2. The pop-up did not block the rest of the page. This design was intentional to avoid rushing students to respond. If a student did not react to the prompt, a pop-up alert was issued every 5 minutes. The question



[1] The *"Organizer"* refers to the assignment front page with links to other sections of the assignment instructions.

Figure 2: The indicators of SRL phases in the context of assignment in the LMS

was multiple-choice (Figure 3, cutout); each choice was designed to describe one of the self-regulatory phases that a student may engage in. The participants were able to choose among the choices or describe their activity. The question asked what they were doing on the assignment at that moment. One choice described: *task definition* ("I determine exactly what the assignment requirements are"), *planning* ("I plan how to do the assignment"), enactment ("I am working on the assignment"), and two choices described: *adapting* ("I identify the remaining tasks to complete the assignment", "I evaluate how far I have progressed in completing the assignment").

2.6 Survey Study

In addition to collecting students' real-time self-reports in Canvas while working on the assignment, we also asked participants to fill out a survey at the end of the study about their self-regulatory behaviors using the self-efficacy and metacognitive self-regulation items of the well-established MSLQ instrument [28, 29]. Among the variety of motivational beliefs involved in self-regulation, self-efficacy has been emphasized [7] since it influences students' goal setting and commitment to those goals, decision making to reach those goals, and their persistence [4]. 32 participants that completed the study filled out the questionnaire. The responses were recorded on a 7-point Likert scale, from 1 (not at all true of me) to 7 (very true of me). The scores for two MSLQ sub-scales, i.e. self-efficacy and self-regulation, were used to create a measure of students' self-regulatory engagement level. Then, agglomerative hierarchical clustering applied to two features identified three groups of students in terms of their self-regulatory engagement levels, i.e. high, moderate, and low self-regulated learners.

Figure 4 shows three clusters of students generated from their responses to the survey. The size of the clusters was reasonably even. High self-regulating students reported that they have a high level of self-efficacy, and they more frequently regulated their learning process (cluster 1, n = 12). The students with moderate (cluster 2, n = 11) and low (cluster 3, n = 9) levels of self-regulatory engagement reported similar self-regulation levels but differed in their self-efficacy level.

3 DATA ANALYSIS

The students' interactions with the Canvas assignment pages (i.e. *Overview, General Guideline, Detail Specification, Resources, Marking Scheme, Submission Instructions*) and related handout files, worked examples, and discussion forum were collected over two weeks as the students worked on each assignment. These streams of click events were segmented into sessions for every student. The cutoff for a session was defined to be 20 minutes, after which subsequent events were considered to be within the next session.

During the classification, we developed an automated system to identify self-regulatory indicators and classify each student's sessions with one of the self-regulatory phases, i.e. *task definition, planning, enactment, and adapting*, which we called **observed SRL tags**. If one or more indicators were found in the session, we applied the macro-ordering of the SRL phases to determine the SRL tag. The list of these indicators is shown in Figure 2. For example, if a student

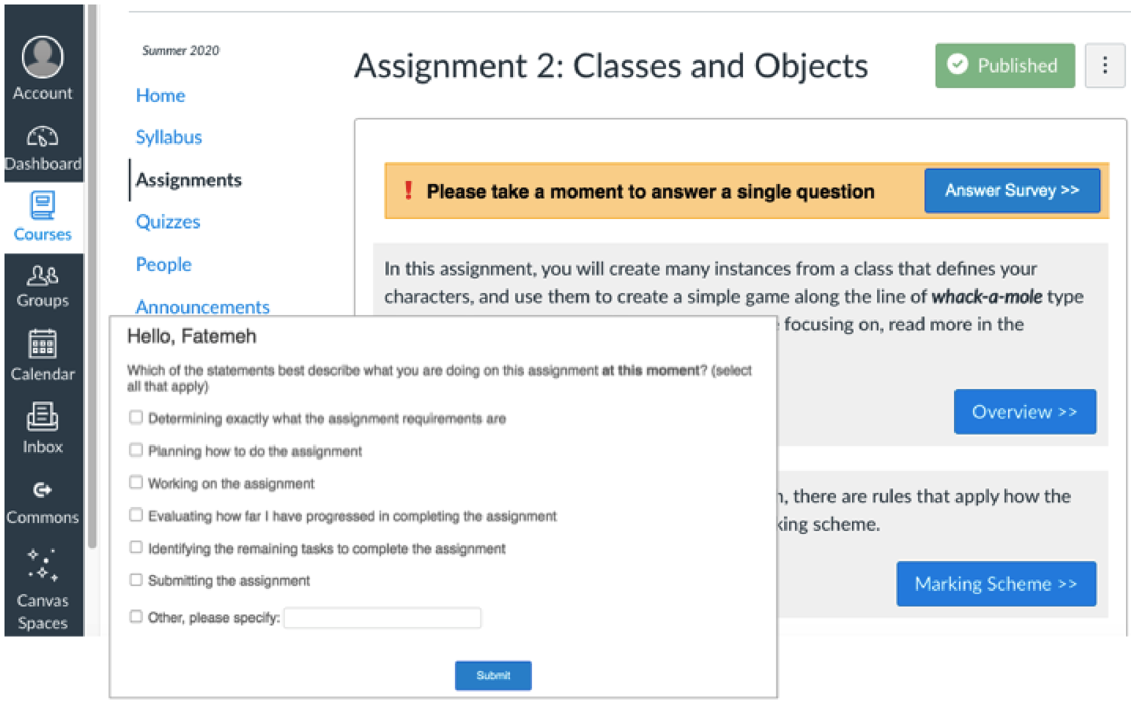


Figure 3: The view of the prompt (orange box) embedded in Canvas and the pop-up question

clicks on (General Guidelines, Overview, Worked Example, Detail Specification) in one session, this session is labeled by *enactment* tag, since it has one of the indicators representing *enactment* phase which is (Worked Example, Detail Specification).

The timestamped self-reported responses were categorized into associated self-regulatory phases, which we called **self-reported SRL tags**. Next, the self-reported and observed SRL tags were merged. The observed SRL tags from Canvas logs were paired with self-reported SRL tags based on the timestamp proximity (nearest neighbor search). Since students were able to stay on Canvas course pages as long as they desired (Canvas did not have a time-out feature set) and our tool would continue prompting students every 20 minutes to answer the question, we had a larger number of self-reported SRL tags; hence we paired the self-reported SRL tag with the observed SRL tags with the closest timestamp.

Finally, to address *RQ3*, we examined the extent to which self-reported and observed SRL tags converged. In this study, two different assignments were analyzed. We calculated weighted Cohen's kappa [10, 33] for each assignment separately to measure corrected for chance agreement between self-reported and observed SRL tags. When calculating kappa, only the perfect classification would count toward an agreement. However, in the weighted kappa statistic, the most weight is given to perfect agreement with less weight given to cells with near-perfect agreement (partial agreement). Because self-regulated learning is a loose cycle of phases that learners engaged in, the weighted kappa provides a better measure of these partial agreements. For instance, the manifestation of *task definition* and *planning* phases are relatively similar in the LMS. It requires detecting the context surrounding student activities to distinguish one from another, and this information is not always available in the LMS, considering the necessarily content-agnostic nature of this learning environment. Moreover, students may not have full awareness or effective control of their metacognition. If a student reported their current behavior as defining a task and our system classified their behavior as a representation of a planning phase, these two classifications would count as a disagreement for the kappa statistic but would count as a partial agreement in the weighted kappa statistic. We calculated and compared both measures in our analyses. Further, we examined the specific sessions which received partial agreement (near perfect agreement) between two classifiers, considering the level of students' self-regulatory engagement measure from the survey study. We discuss the results in the next section.

4 RESULTS

4.1 Comparing Classifications in Assignment 2

The total number of 243 sessions in Assignment 2 was identified for 38 students. For each session, the participants reported what they were doing on the assignment (self-reported SRL tags) and our classifier labeled that session based on the detected indicators of SRL phases in the log data (observed SRL tags). The results of classifications in Assignment 2 are shown in Table 1, where the values are reported in terms of frequencies and percentages (e.g., the value of the first cell Task Definition indicates that 37 out of 243 sessions were classified as the students were in *task definition* phase). The marginal row totals indicate the percentage of observed SRL tags by our behavioral classifier. Similarly, the column totals indicate the

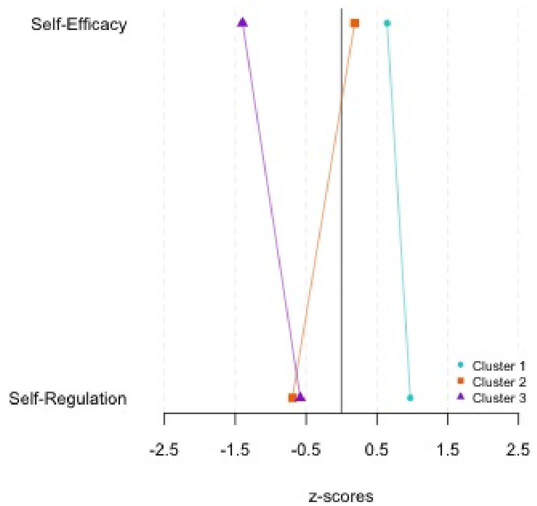


Figure 4: Participants' self-regulatory groups based on survey study

number of self-reported SRL tags by students. In Assignment 2, the students self-reported 19% of total sessions as *task definition* phase, and our system placed 22% of the sessions in that phase. Similarly, the students reported 17% of total sessions as *planning* and the system classified 20% in this phase. The students reported 37% as *enactment* and 27% as *adapting*, whereas the system labeled 18% as enactment and 40% as *adapting*. The diagonal values of Table 1 represent classifications on which the two SRL tags agreed exactly, which indicates that perfect agreement was reached on 61% of the sessions. However, the chance-corrected agreement is moderate (κ = *.48*). Since the largest confusion is between classifying *enactment* and *adapting* phases, and these two phases are adjacent in the SRL cycle, we calculated the quadratic weighted Cohen's kappa to account for partial agreement (e.g. enactment and adapting) between these two sets of tags. The result showed substantial agreement [21] between self-reported SRL tags and observed SRL tags (κ = *.74*).

4.1.1 Examining partial agreements on session classification. Since survey data was not complete for all students in Assignment 2, we identified a total of 224 sessions for 32 students. As noted above, the largest percentage of confusion in Table 1 was when the students reported *enactment* while the automated classifier categorized those sessions as *adapting* phase. Another noticeable confusion was when the students reported *enactment* and our classifier labeled them as *planning*. We examined how differences in students' self-regulatory engagement levels influenced these discrepancies. We examined 85 sessions that were reported as *enactment* by students. The total number of sessions was different from what is shown in Table 1 due to the reduced number of students who participated in the survey study at the end of the study. As shown in Table 2, all three groups of students (high, moderate, and low self-regulatory learners) reached a perfect agreement with the automated classifier on *enactment* almost equally. The highest number of discrepancies (17 out of 85) belongs to the high self-regulated group when they reported *enactment* and our classifier categorized these same sessions as *adapting*.

4.2 Comparing Classifications in Assignment 3

We recorded the total number of 187 sessions in Assignment 3 from 32 participants. The results are shown in Table 3. Perfect agreement between observed and self-reported SRL tags was found in 59% of sessions. The chance-corrected agreement was moderate (κ = *.43*). The students reported 22% as *task definition* phase, and 10% as *planning* phase. On the contrary, our classifier categorized 14% of total sessions as *task definition* and 22% as *planning*. The participants reported *enactment* and *adapting* in 44% and 24% of sessions, respectively. Our system placed 33% of the sessions in *enactment* and 30% in *adapting* phase. Similarly, we calculated quadratic weighted Cohen's kappa coefficient to account for partial agreement, since again the highest percentage of confusion was between *planning, enactment* and *adapting, enactment*, which are adjacent phases in the SRL cycle. The self-reported and observed SRL classifications reached partial agreement, which was verified by weighted Cohen's kappa (κ = *.68*).

4.2.1 Examining partial agreements on session classification. We examined further the sessions having high non-identical tags. Similar to Assignment 2, the highest disagreement belongs to sessions classified as *adapting* whereas the students reported them as *enactment* phase. In addition, there are sessions which were categorized as *planning* by the automated classifier while the students equally reported them as either *task definition* or *enactment*. Table 4 shows how three groups of self-regulated learners reported these sessions as *enactment* whereas these same sessions were classified mostly as either *adapting* or *planning* by our classifier. Similar to Assignment 2, the high self-regulated learners had the highest level of disagreement with the automated classifier (14 out of 83) comparing to their peers.

Given relatively equal distributions of disagreement among three self-regulatory groups in two assignments, as shown in Tables 2 & 4, we examined whether there is an association between students' reporting of *enactment* and the observed tags of the sessions in two assignments. The association was not significant (χ^2 = *11.48, p = .07*).

5 DISCUSSION

The present study contributes to research on modeling students' self-regulated learning in an open-ended learning environment (LMS). We created visible markers of student learning progress at different points in time in the LMS by utilizing instructional design, which can be reproduced in various courses and disciplines in Canvas. We adopted a theoretically-based and empirically-derived approach to collecting and modeling multi-channel data to extend the current SRL frameworks. Focusing on an open-ended technology (LMS), which is used widely across higher education, we framed our study around three research questions to develop indicators of self-regulatory phases in the context of the problem-solving task.

Regarding the first research question, we developed proximal indicators of self-regulated learning phases using trace data (LMS logs) of students' interactions with the LMS while working on the

Table 1: Observed SRL tags vs. students' self-reported SRL tags for Assignment 2

	Students' Self-Reported SRL Tag				
Observed SRL Tag	Task Definition	Planning	Enactment	Adapting	Total
Task Definition	37 (15%)	10 (4%)	5 (2%)	1 (0%)	53 (22%)
Planning	4 (2%)	24 (10%)	20 (8%)	1 (0%)	49 (20%)
Enactment	2 (1%)	3 (1%)	32 (13%)	7 (3%)	44 (18%)
Adapting	3 (1%)	4 (2%)	34 (14%)	56 (23%)	97 (40%)
Total	46 (19%)	41 (17%)	91 (37%)	65 (27%)	243 (100%)

	Students' SRL level		
Observed SRL Tag	High	Moderate	Low
Task Definition	2 (6%)	1 (4%)	1 (4%)
Planning	7 (19%)	3 (13%)	7 (27%)
Enactment	10 (28%)	11 (48%)	10 (38%)
Adapting	17 (47%)	8 (35%)	8 (31%)

Table 2: The observed tags which were reported as *enactment* among three levels of SRL groups (Assignment 2)

assigned tasks. We addressed three major challenges to detect self-regulatory phases in LMS logs. First was identifying the underlying patterns of self-regulatory phases to develop the proximal indicators from log data. We adopted a theoretically-based approach to develop these indicators from a sample of authentic LMS logs. Second, creating the representations of the theoretical assumptions from trace data was challenging. For instance, to do a download of worked example and revisit of assignment instructions represented an *enactment* phase in the process of a problem-solving task. We used multi-channel data as a source of ground truth for behavior detection. Another challenge was the granularity level at which self-regulatory phases should be detected. We proposed to create a macro-level sequence or what we called "a sequence of sessions" to account for the underlying temporality of the self-regulated learning process.

Concerning the second research question, we developed a prompting tool, embedded in the LMS assignment pages, to collect students' real-time self-report behaviors in terms of four self-regulatory phases. The prompting schedule was designed based on the granularity level at which the indicators of self-regulatory phases are found in the log data. We conducted a quasi-experiment to collect real-time self-reports in two consecutive assignments. The main challenge in this phase of the study was the temporal alignment of multi-channel SRL data based on different sampling rates. The number of times that students reported their behaviors was more frequent than the logs' sessions because LMS (Canvas) only stored click events initializing the opening of the pages. However, students could stay in the assignment pages longer than the session cutoff and be asked to report their behaviors again. We aligned these two data sets using their closest timestamps. The timestamps were not always exactly aligned. We set the threshold of 20 minutes of absolute difference between the timestamps of self-reported tags and corresponding observed tags (M = 4m 05s, SD = 3m 40s). Our findings indicate that, in particular, the self-reported SRL tags of *enactment* phase did not converge with their corresponding observed SRL tags in ~20% of the sessions. Research [11] also supported our findings that students may evaluate different sub-components of the factors when they compare multiple measures of metacognition (e.g. planning and monitoring correlated to evaluation and reflection). Furthermore, we were concerned with not overburdening students with the number of times they should report their behaviors, so we used a single item survey. A few students used the "Other" option frequently and wrote their responses. Some of the answers detailed what they were doing on the assignment (e.g. "I'm really stuck- I can't complete Sketch 3, go to bed."). The willingness to provide additional information made us consider revising the survey in future studies to collect more information about how students exercise self-regulated learning.

Finally, to address the third research question about the extent to which behavioral indicators converge with self-report SRL phases, we computed Cohen's kappa statistic on the 243 sessions in the second assignment and 187 sessions in the third assignment. The results shown in Tables 1 & 3 indicate substantial agreement if the most weight is given to perfect agreement (dark gray) with less weight given to cells with near-perfect agreement (light gray). Specifically, self-report and automated classifier did not exactly agree on *enactment* and *adapting* phases. In the second assignment, the students reported 14% of the sessions as *enactment*, which were labeled as *adapting* by the automated classifier. The disagreement was fairly similar in the third assignment. The students reported 11% of sessions as *enactment*, which were classified as its adjacent phase *adapting*. We investigated these sessions further, to assess which students reported *enactment* while our classifier categorized these same sessions as either *planning* or *adapting* phases. The high self-regulated students tended to disagree with the behavioral classifier more than other groups on *adapting* (31 out of 168 sessions collectively). They reported these sessions as *enactment*. The results were not consistent for those sessions, which were classified as *planning*. In particular, there is not a significant association among students with different levels of self-regulation and the observed SRL tags of those sessions reported as *enactment* phase. The results suggest that our indicators are comparably successful at classifying SRL phases for different self-regulatory engagement levels. Next, we examined which indicators led to partial agreements of these sessions. The indicator (Organizer, Overview, General Guideline, Detail Specification), is a common representation of *planning* and *enactment* phases. In addition, the indicator (Organizer, Detail Specification) is shared by *enactment* and *adapting* phases. In these cases, the macro-level sequence, i.e. prior session classification, was a deterministic factor to identify the observed SRL tag. For instance,

Table 3: Observed SRL tags vs. students' self-reported SRL tags for Assignment 3

Observed SRL Tag	Students' self-reported SRL Tag: Task Definition	Planning	Enactment	Adapting	Total
Task Definition	19 (10%)	3 (2%)	4 (2%)	1 (.5%)	27 (14%)
Planning	14 (7.5%)	13 (7%)	14 (7.5%)	0 (0%)	41 (22%)
Enactment	6 (3%)	3 (2%)	44 (24%)	9 (5%)	62 (33%)
Adapting	2 (1%)	0 (0%)	21 (11%)	34 (18%)	57 (30%)
Total	41 (22%)	19 (10%)	83 (44%)	44 (24%)	187 (100%)

Observed SRL Tag	Students' SRL level: High	Moderate	Low
Task Definition	0 (0%)	0 (0%)	4 (12%)
Planning	4 (12%)	4 (23%)	6 (18%)
Enactment	15 (47%)	11 (65%)	18 (55%)
Adapting	14 (41%)	2 (12%)	5 (16%)

Table 4: The observed tags which were reported as *enactment* among three levels of SRL groups (Assignment 3)

if there is an indicator of (Organizer, Detail Specification) within a session and its prior session is *enactment*, then the session is classified as *adapting* phase. However, our results suggest that distinguishing *planning* and *adapting* phases from an adjacent phase, such as *enactment*, requires more information.

Overall, our study showed that the instructional design of learning activity in the LMS could support learning analytics to uncover the context of students' learning behaviors, such as their progress through the SRL phases. We have developed SRL indicators from the students' log data while interacting with the LMS in the context of the problem-solving process. We examined how the indicators, after being applied them to the students' clickstream in the next offering of the same course with two different activities, aligned with students' self-reported behaviors. Relatively high weighted Cohen's kappa agreements indicated the robustness of indicators in detecting student SRL phases in computer programming tasks of similar scope. The similar results from two subsequent assignments suggest that the indicators are, to some extent, resilient to the learning effect where students may use specific information more than others (e.g. Detail Specification section of instructions). Furthermore, our study addressed some of the important challenges in modeling SRL. In particular, in open-ended technologies like an LMS, in which students also engage in learning activities within the intervals between events observed in the learning environment (so-called censored observations), the triangulation of different sources of students' data benefits not only the cultivation of data-driven decision-making but also our understanding of how they perceive their own learning experience outside of the content provided by instructors.

A further critical aspect of detecting SRL behaviors concerns measuring the qualities of students' SRL. In this particular study, we strived to create representations of SRL phases and examine whether they matched with what students reported. As SRL is a complex metacognitive process, it is not directly observable. Thus, measuring the qualities of self-regulated learning in each phase is difficult to capture at scale in online learning environments, but it could significantly improve practice and consequently help learners to learn more effectively. Winne [41] stated that each time learners engage in SRL process, they potentially do an experiment by gathering and analyzing data about why their approaches to learning are more or less successful. He recommended educators and instructional designers offer heuristics for practices instead of providing what-to-do instructions. If we move forward to learning design approach that promotes the learner being in charge, we may have richer trace data that help further to measure the qualities of students' engaging in SRL phases.

5.1 Study Limitations and Future Directions

The present work is the first empirical study to examine our framework for identifying self-regulatory phases in an LMS. One limitation was that Canvas log tracing capacity did not record how long students stayed on an individual page. Thus click event logs resulted in a loss of part of our self-reported data in the process of merging with Canvas log data. Consequently, some self-reported tags were aligned with the same observed SRL tags that had the closest timestamps. This may cause a partial disagreement between self-reported and observed tags found in this study. Finally, as with all studies where participation is voluntary, the results may be influenced by the self-selection of participants.

Our immediate next steps are already in progress. Following the recent work by [25], we developed a plug-in to log duration-based events in Canvas. This means we can track which Canvas page students attend and for how long. These duration-based log data will allow us to have a more precise alignment between self-reported and observed SRL tags. We follow a similar research protocol and apply our framework to more diverse learning activities across university courses. Our assignment design template has already been used in other courses that targeted different cognitive skills such as critical thinking. In the current semester, we follow a similar approach to collecting students' real-time self-reported behaviors, and their trace data involving 350 students in 2 large undergraduate courses. We intend to improve our framework robustness, in particular, to differentiate adjacent phases of *enactment* and *adapting*.

From the outset, we have designed the assignment template with a vision to expand on our campus. Given the extensive penetration of Canvas in the LMS marketplace, such modification could have a significant impact. As current empirical research [1] on learning indicate that undergraduate students have difficulty learning about problem-solving, reasoning, and more complex topics in STEM areas and shows that metacognitive skills are teachable [26], this framework of self-regulatory phases may have the potential to provide real-time adaptive scaffolding and/or feedback for students at scale with different self-regulatory needs.

ACKNOWLEDGMENTS

We wish to acknowledge the support of the Natural Sciences and Engineering Research Council of Canada (NSERC), RGPIN-2018-06071, and Social Sciences and Humanities Research Council (SSHRC), 435-2016-1205. We would also like to thank Dr. Philip Winne and the late Dr. Stuart Karabenick for their insight and guidance to support the first author in this project.

REFERENCES

[1] Roger Azevedo and Dragan Gašević. 2019. Analyzing multimodal multichannel data about self-regulated learning with advanced learning technologies: Issues and challenges.
[2] Roger Azevedo, Michelle Taub, Nicholas V Mudrick, SA Martin, and J Grafsgaard. 2018. Using multi-channel trace data to infer and foster self-regulated learning between humans and advanced learning technologies. *Handbook of self-regulation of learning and performance* 2 (2018).
[3] Ryan SJD Baker, Albert T Corbett, and Angela Z Wagner. 2006. Human classification of low-fidelity replays of student actions. In *Proceedings of the Educational Data Mining Workshop at the 8th International Conference on Intelligent Tutoring Systems*, Vol. 2002. 29–36.
[4] Albert Bandura. 1993. Perceived self-efficacy in cognitive development and functioning. *Educational psychologist* 28, 2 (1993), 117–148.
[5] Maria Bannert, Peter Reimann, and Christoph Sonnenberg. 2014. Process mining techniques for analysing patterns and strategies in students' self-regulated learning. *Metacognition and learning* 9, 2 (2014), 161–185.
[6] Sanam Shirazi Beheshitha, Dragan Gašević, and Marek Hatala. 2015. A process mining approach to linking the study of aptitude and event facets of self-regulated learning. In *Proceedings of the fifth international conference on learning analytics and knowledge*. 265–269.
[7] Deborah L Butler and Philip H Winne. 1995. Feedback and self-regulated learning: A theoretical synthesis. *Review of educational research* 65, 3 (1995), 245–281.
[8] Analía Cicchinelli, Eduardo Veas, Abelardo Pardo, Viktoria Pammer-Schindler, Angela Fessl, Carla Barreiros, and Stefanie Lindstädt. 2018. Finding traces of self-regulated learning in activity streams. In *Proceedings of the 8th International Conference on Learning Analytics and Knowledge*. ACM, 191–200.
[9] Timothy J Cleary, Gregory L Callan, Jaime Malatesta, and Tanya Adams. 2015. Examining the level of convergence among self-regulated learning microanalytic processes, achievement, and a self-report questionnaire. *Journal of Psychoeducational Assessment* 33, 5 (2015), 439–450.
[10] Jacob Cohen. 1968. Weighted kappa: nominal scale agreement provision for scaled disagreement or partial credit. *Psychological bulletin* 70, 4 (1968), 213–220.
[11] Kym Craig, Daniel Hale, Catherine Grainger, and Mary E Stewart. 2020. Evaluating metacognitive self-reports: systematic reviews of the value of self-report in metacognitive research. *Metacognition and Learning* 15, 2 (2020), 155–213.
[12] K Anders Ericsson and Herbert A Simon. 1984. *Protocol analysis: Verbal reports as data.* The MIT Press.
[13] Dragan Gašević, Shane Dawson, Tim Rogers, and Danijela Gasevic. 2016. Learning analytics should not promote one size fits all: The effects of instructional conditions in predicting academic success. *The Internet and Higher Education* 28 (2016), 68–84.
[14] Dragan Gasevic, Jelena Jovanovic, Abelardo Pardo, and Shane Dawson. 2017. Detecting learning strategies with analytics: Links with self-reported measures and academic performance. *Journal of Learning Analytics* 4, 2 (2017), 113–128.
[15] Jeffrey A Greene and Roger Azevedo. 2010. The measurement of learners' self-regulated cognitive and metacognitive processes while using computer-based learning environments. *Educational psychologist* 45, 4 (2010), 203–209.
[16] Jeffrey A Greene and Dale H Schunk. 2017. Historical, contemporary, and future perspectives on self-regulated learning and performance. In *Handbook of Self-Regulation of Learning and Performance*. Routledge, 17–32.
[17] Allyson F Hadwin, John C Nesbit, Dianne Jamieson-Noel, Jillianne Code, and Philip H Winne. 2007. Examining trace data to explore self-regulated learning. *Metacognition and Learning* 2, 2-3 (2007), 107–124.
[18] Dirk Ifenthaler. 2012. Determining the effectiveness of prompts for self-regulated learning in problem-solving scenarios. *Journal of Educational Technology & Society* 15, 1 (2012), 38–52.
[19] Jelena Jovanović, Dragan Gašević, Shane Dawson, Abelardo Pardo, Negin Mirriahi, et al. 2017. Learning analytics to unveil learning strategies in a flipped classroom. *The Internet and Higher Education* 33, 4 (2017), 74–85.
[20] Stuart A Karabenick and Akane Zusho. 2015. Examining approaches to research on self-regulated learning: conceptual and methodological considerations. *Metacognition and Learning* 10, 1 (2015), 151–163.
[21] J Richard Landis and Gary G Koch. 1977. The measurement of observer agreement for categorical data. *biometrics* (1977), 159–174.
[22] Steven Lonn, Stephen Aguilar, and Stephanie D Teasley. 2013. Issues, challenges, and lessons learned when scaling up a learning analytics intervention. In *Proceedings of the third international conference on learning analytics and knowledge*. ACM, 235–239.
[23] Jonna Malmberg, Sanna Järvelä, Jukka Holappa, Eetu Haataja, Xiaohua Huang, and Antti Siipo. 2019. Going beyond what is visible: What multichannel data can reveal about interaction in the context of collaborative learning? *Computers in Human Behavior* 96 (2019), 235–245.
[24] Wannisa Matcha, Dragan Gašević, Jelena Jovanović, Abelardo Pardo, Jorge Maldonado-Mahauad, Mar Pérez-Sanagustín, et al. 2019. Detection of Learning Strategies: A Comparison of Process, Sequence and Network Analytic Approaches. In *European Conference on Technology Enhanced Learning*. Springer, 525–540.
[25] Nicholas V Mudrick, Roger Azevedo, and Michelle Taub. 2019. Integrating metacognitive judgments and eye movements using sequential pattern mining to understand processes underlying multimedia learning. *Computers in Human Behavior* 96 (2019), 223–234.
[26] John Perry, David Lundie, and Gill Golder. 2019. Metacognition in schools: what does the literature suggest about the effectiveness of teaching metacognition in schools? *Educational Review* 71, 4 (2019), 483–500.
[27] Paul R Pintrich. 1995. Understanding self-regulated learning. *New directions for teaching and learning* 1995, 63 (1995), 3–12.
[28] Paul R Pintrich and Elisabeth V De Groot. 1990. Motivational and self-regulated learning components of classroom academic performance. *Journal of educational psychology* 82, 1 (1990), 33–40.
[29] Paul R Pintrich, David AF Smith, Teresa Garcia, and Wilbert J McKeachie. 1993. Reliability and predictive validity of the Motivated Strategies for Learning Questionnaire (MSLQ). *Educational and psychological measurement* 53, 3 (1993), 801–813.
[30] Bart Rienties, Lisette Toetenel, and Annie Bryan. 2015. Scaling up learning design: impact of learning design activities on lms behavior and performance. In *Proceedings of the Fifth International Conference on Learning Analytics And Knowledge*. ACM, 315–319.
[31] Ido Roll and Philip H Winne. 2015. Understanding, evaluating, and supporting self-regulated learning using learning analytics. *Journal of Learning Analytics* 2, 1 (2015), 7–12.
[32] Sanne FE Rovers, Geraldine Clarebout, Hans HCM Savelberg, Anique BH de Bruin, and Jeroen JG van Merriënboer. 2019. Granularity matters: comparing different ways of measuring self-regulated learning. *Metacognition and Learning* 14, 1 (2019), 1–19.
[33] Neil J Salkind. 2006. *Encyclopedia of measurement and statistics*. SAGE publications.
[34] Michael A Sao Pedro, Ryan SJD Baker, and Janice D Gobert. 2013. What different kinds of stratification can reveal about the generalizability of data-mined skill assessment models. In *Proceedings of the Third International Conference on Learning Analytics and Knowledge*. 190–194.
[35] Benjamin Saunders, Julius Sim, Tom Kingstone, Shula Baker, Jackie Waterfield, Bernadette Bartlam, Heather Burroughs, and Clare Jinks. 2018. Saturation in qualitative research: exploring its conceptualization and operationalization. *Quality & quantity* 52, 4 (2018), 1893–1907.
[36] John R Savery. 2015. Overview of problem-based learning: Definitions and distinctions. *Essential readings in problem-based learning: Exploring and extending the legacy of Howard S. Barrows* 9 (2015), 5–15.
[37] Gayane Sedrakyan, Jonna Malmberg, Katrien Verbert, Sanna Järvelä, and Paul A Kirschner. 2018. Linking learning behavior analytics and learning science concepts: designing a learning analytics dashboard for feedback to support learning regulation. *Computers in Human Behavior* (2018).
[38] James R Segedy, John S Kinnebrew, and Gautam Biswas. 2015. Coherence over time: understanding day-to-day changes in students' open-ended problem solving behaviors. In *International Conference on Artificial Intelligence in Education*. Springer, 449–458.
[39] James R Segedy, John S Kinnebrew, and Gautam Biswas. 2015. Using Coherence Analysis to Characterize Self-Regulated Learning Behaviours in Open-Ended Learning Environments. *Journal of Learning Analytics* 2, 1 (2015), 13–48.
[40] Christoph Sonnenberg and Maria Bannert. 2019. Using Process Mining to examine the sustainability of instructional support: How stable are the effects of metacognitive prompting on self-regulatory behavior? *Computers in Human Behavior* 96 (2019), 259–272.
[41] Philip H Winne. 2018. Cognition and metacognition within self-regulated learning. (2018).
[42] Philip H Winne and Allyson F Hadwin. 1998. Studying as self-regulated engagement in learning. *Metacognition in educational theory and practice* (1998), 277–304.

[43] Philip H Winne and Dianne Jamieson-Noel. 2002. Exploring students' calibration of self reports about study tactics and achievement. *Contemporary Educational Psychology* 27, 4 (2002), 551–572.

[44] Mingming Zhou and Philip H Winne. 2012. Modeling academic achievement by self-reported versus traced goal orientation. *Learning and Instruction* 22, 6 (2012), 413–419.