JOURNAL OF LEARNING ANALYTICS

Volume 12(3), 66–86. https://doi.org/10.18608/jla.2025.8865

Interpreting Predictive Learning Sequences in a College Math Course through a Self-Regulated Learning Framework

Linyu Yu[1], Peter F. Halpin[2], Matthew L. Bernacki[3], Sirui Ren[4], Robert D. Plumley[5] and Jeffrey A. Greene[6]

Abstract

Digital traces have been used to measure self-regulated learning (SRL), yet the validity of inferences made about these traces has often been questioned. Recently, researchers have used multiple channels of data — including digital traces, verbalizations, and self-reports — to validate inferences about individual SRL events. Research on the validation of inferences about sequences of multiple SRL events remains limited; however, investigating these sequences has the potential to refine SRL theories. To study the validation of sequences of SRL events, we collected multimodal data from 49 undergraduates completing a math task in a lab setting. Participants were asked to think aloud while interacting with different digital platforms. Then, we used sequence pattern mining to identify the digital events most predictive of post-test scores. Next, we used student verbalizations during the learning process to validate the inferences about what those predictive sequences reflected. Sequences representing learner conscientiousness predicted better performance; sequences that included pausing and rewinding videos predicted poorer performance. Some learner verbalizations co-occurred with digital events and consistently aligned with SRL processes, providing validity evidence for SRL sequences. Heterogeneity in verbal-to-digital trace alignment emerged and will require methodological advances to validate the sequences specific to individuals and task conditions.

Notes for Practice

- Student verbalizations provide useful data to understand self-regulated learning (SRL) processes.
- The alignment between digital trace data and learner verbalization allowed us to provide a theory-driven interpretation of trace data in terms of the SRL processes that they reflect.
- Video-watching sequences that involve pausing and rewinding were negatively associated with performance. Learning designs should prioritize capturing digital traces that reflect student engagement in active learning environments.

Keywords: Self-regulated learning, digital events, sequence mining, think-aloud verbalizations, validation

Submitted: 13/01/2025 — **Accepted:** 18/07/2025 — **Published:** 26/11/2025

Corresponding author [1]Email: linyu@live.unc.edu Address: The University of North Carolina at Chapel Hill, CB 3500 Peabody Hall, Chapel Hill, NC 27599-3500, USA. ORCID iD: https://orcid.org/0000-0002-2021-8811

[2]Email: peter.halpin@unc.edu Address: The University of North Carolina at Chapel Hill, CB 3500 Peabody Hall, Chapel Hill, NC 27599-3500, USA. ORCID iD: https://orcid.org/0000-0002-2519-2841

[3]Email: mlb@unc.edu Address: The University of North Carolina at Chapel Hill, CB 3500 Peabody Hall, Chapel Hill, NC 27599-3500, USA; Brain & Motivation Research Institute, Department of Education, Korea University, 145 Anam-ro, Seongbuk-gu, Seoul 02841, Korea. ORCID iD: https://orcid.org/0000-0003-1279-2829

[4]Email: rensirui@unc.edu Address: The University of North Carolina at Chapel Hill, CB 3500 Peabody Hall, Chapel Hill, NC 27599-3500, USA. ORCID iD: https://orcid.org/0009-0008-8480-8459

[5]Email: rplumley@live.unc.edu Address: The University of North Carolina at Chapel Hill, CB 3500 Peabody Hall, Chapel Hill, NC 27599-3500, USA. ORCID iD: https://orcid.org/0000-0001-8979-6276

[6]Email: jagreene@email.unc.edu Address: The University of North Carolina at Chapel Hill, CB 3500 Peabody Hall, Chapel Hill, NC 27599-3500, USA. ORCID iD: https://orcid.org/0000-0003-4145-1847

1. Introduction

College courses like Introductory Math often act as gatekeepers for undergraduates' pursuit of science, technology, engineering, or math (STEM) majors and careers, and many such courses have transitioned from lecture to active learning formats (Eddy & Hogan, 2014; Olson & Riordan, 2012). Active learning approaches have been found to increase student engagement and critical thinking skills and improve their overall academic performance (Eddy & Hogan, 2014; Theobald et al., 2020). Courses that incorporate these approaches can be challenging for students, however, since they often require self-motivation and effortful self-regulated learning (SRL; Bernacki et al., 2025; Shekhar et al., 2020). Few students receive instruction in either of these skills prior to enrolment (Sebesta & Speth, 2017).

SRL is the process by which individuals set goals for their learning, monitor their progress, and actively control their cognition, metacognition, motivation, affect, behaviours, and environment to achieve those goals (Greene et al., 2024). SRL is strongly correlated with academic success across educational levels (Dent & Koenka, 2016; Sebesta & Speth, 2017). Given its strong link to academic outcomes, researchers have explored various methods to study SRL. Many researchers have studied it in laboratory settings using think-aloud protocols (e.g., Bannert, 2007; Greene et al., 2018). Increasingly, researchers have begun to analyze logs of digitally traced SRL events that students produce when engaging in learning tasks using technologies (Bernacki, 2018; Greene & Azevedo, 2010; Winne, 2017, 2022). These digital traces provide key information about student engagement with the learning materials, such as access time, activity, duration, number of attempts, and correctness of attempts, among others (Bernacki, 2018). The shift toward analyzing digital traces allows researchers to study SRL in more naturalistic and scalable ways (Bernacki et al., 2025). However, the convenience of learning analytics methods also comes with a drawback; the data often only shows *what* learners did, rather than *why* they adopted these learning processes (Siemens, 2013). Thus, researchers often struggle to confidently infer the specific learning processes reflected in these digital events (Wise & Shaffer, 2015).

In recent years, researchers have also begun to examine sequences of digital traces, due to their potential to capture more complex learning processes than what can be modelled using individual traces and also their potential to examine the dynamic and adaptive nature of SRL over time (e.g., Nath et al., 2024). Some have adopted a theory-driven approach, drawing on existing literature to interpret the meaning of action sequences. For example, Maldonado-Mahauad et al. (2018) categorized student behaviours in a MOOC into six types of learning actions and used process mining to identify frequent sequences. These were then mapped to SRL processes such as elaboration, evaluation, help-seeking, and task exploration based on prior research. Others have combined this theory-driven approach with a data-driven strategy by incorporating additional data sources. Fan, van der Graaf, et al. (2022), for instance, compared theory-driven coding with student verbalizations, using the verbal data to refine the interpretation of sequences. A critical step in this process is trace parsing, determining which actions mark the beginning and end of a meaningful learning process. Recent studies (e.g., Osakwe et al., 2024) have shown that different trace parsing methods can yield different sequences, and when aligned with verbalizations, can lead to different interpretations of SRL processes. One way to address this challenge is to first identify a manageable set of meaningful sequences and then use multiple data channels to interpret them more accurately.

In this study, we investigated SRL processes that undergraduates enacted within a flipped introductory algebra course, where content delivery occurred outside of class through asynchronous video lectures and practice problems; class time was dedicated to active practice and discussions (Bishop & Verleger, 2013). Specifically, we had two primary aims:

1. To examine how sequences of digital traces enacted by undergraduate students during flipped lessons predict their achievement in instructional tasks.
2. To determine whether we could use verbalizations to understand and validate inferences about the SRL processes indicated by sequences of digital traces.

In addressing the first aim, we sought to provide pragmatic insights into the specific behaviours and event patterns that characterize successful engagement with digital resources in flipped classroom settings. In addressing the second aim, we focused on aligning student verbalizations with their digital event sequences to offer a theoretical lens for interpreting SRL processes (Winne & Hadwin, 2008). These insights can guide instructors in designing lessons, structuring tasks, and creating digital content that better supports students in productively interacting with materials to achieve their learning goals.

2. Literature Review

2.1. Self-Regulated Learning (SRL) Processes and Measures to Observe Them

Early SRL research relied on classroom observation (Zimmerman, 1986) and self-reports (e.g., Motivated Strategies for Learning Questionnaire; Pintrich et al., 1991). Given concerns about the accuracy and specificity of such measures, researchers have tended toward event-based measures of SRL including the use of think-aloud protocols (Greene et al., 2018). Learner verbalizations are then coded into cognitive, affective, metacognitive, and motivational subprocesses (Winne et al., 2002). For example, if a learner says, "*I don't think I learned that concept well*," this verbalization will be coded as a negative judgment of learning within the *metacognitive monitoring* category.

Verbalization data are often considered more accurate indicators of SRL than self-report questionnaires, which are more likely to be biased by memory loss and recollection issues (Russell & Winston, 2014). However, due to the time and effort required to invite learners to the lab, train them to verbalize, and train and deploy researchers to code those verbalizations, think-aloud studies are often limited in sample size and thus limited in their generalizability (Hu & Gao, 2017; Zhang et al., 2024).

Increasingly, researchers have used *digital traces* collected unobtrusively in computerized learning environments as indicators of SRL (Greene & Azevedo, 2010) to provide rich data about the time a learner engages with a resource, the action they undertake, and the results of that action. We can use this information to make inferences about how the resources are meant to afford cognitive and metacognitive SRL processes (Bernacki, 2018). However, understanding *why* individual learners engaged with resources, and whether we should make the same inference across multiple learners remains unclear. Calls to validate such assumptions have emerged in the SRL research community (Winne, 2020). Multimodal data — including language analysis, eye-tracking, video, and biosensors — provide an opportunity to examine whether data from multiple modalities overlap and corroborate each other (Blikstein & Worsley, 2016; Giannakos et al., 2022).

2.2. Validation of Inferences Made From Digital Trace Data Using Multimodal Data

Validity refers to the extent to which empirical evidence and theoretical reasoning support the accuracy and appropriateness of inferences drawn from data (Messick, 1989). In studying SRL, validation efforts provide confidence in the references made about digital trace data as meaningful reflections of SRL processes. In recent validation studies, researchers have adopted multimodal approaches, integrating multiple data sources to understand the inferences made about digital traces. For example, Fan, Lim, et al. (2022) validated the measurement of digital trace data by aligning channels of peripheral (e.g., mouse clicks and keyboard strokes) and eye-tracking data. They found that adding multiple channels allowed researchers to detect more complex SRL processes than using one data channel. For example, an action labelled as *Reading* was reinterpreted as *Organization* — a more complex SRL process — when an additional data channel was added to provide richer context about the action.

One limitation in Fan, Lim et al.'s study (2022) was that they relied on tasks and environments specifically designed by researchers, which may not fully represent authentic learning scenarios. To address this concern, other researchers have embedded validation measures within learning contexts designed by course instructors, aiming to capture SRL processes as they naturally occur in real educational settings. For example, Salehian Kia et al. (2021) validated traces in an authentic learning environment by embedding pop-up questions in Canvas (the learning management system used in the course) during a two-week programming task. They used Cohen's kappa to assess the alignment between digital traces and self-reported SRL phases based on Winne and Hadwin's (1998) model.

Salehian Kia's (2021) study offers ecological validity, but new concerns arise when interpretations of SRL phase labels are made by students themselves. As Karabenick et al. (2007) highlighted, learners may interpret prompts differently, leading to varying SRL phase labels for similar tasks. Bernacki et al. (2025) aimed to overcome issues of authenticity and interpretation by having learners think-aloud in a lab session when engaging with a lesson sampled from a course where all participants were already enrolled. The researchers validated the inferences of digital traces by aligning them with verbalized SRL processes within a 20-second window, meaning using verbalizations that happened 10 seconds before or after a digital trace to infer the meaning of the digital trace. They found that more than 80% of digital traces captured in their study aligned temporally with overlapping verbalizations. Of the 14 types of digital events that could be generated during the task, 10 traces were matched with a modal verbal SRL code. When these 10 varieties of traced events were observed in classroom settings, student engagement in the validated SRL processes were predictive of their performance on exams and course grades, as theorized.

Across all these efforts involving validation via peripheral, eye-tracking, self-report, and verbal data channels, SRL processes have been inferred from *single events*. These single events fail to fully capture nuanced, contextualized processes (Winne, 2020) because they do not entirely account for the temporal and sequential nature of learning as it unfolds in real-world contexts (Ben-Eliyahu & Bernacki, 2015). Sequences of events have been found to explain greater amounts of the variance in performance than single events (Malekian et al., 2020). Their superior *a priori* representation of SRL and predictive power highlight the need for researchers to move beyond single-event analysis to sequential analysis, and validation of emergent, sequence-based SRL processes. To do so, the first step is identifying sequences of interest. Sequence pattern mining (SPM) is one promising approach to explore these sequences.

2.3. Sequence Pattern Mining

SPM was developed to find interesting sequences in a dataset (Dong & Pei, 2007; Zhang & Paquette, 2023). In educational research, a primary goal of SPM studies is to find sequences linked to varying levels of academic performance (Bakhshinategh et al., 2018). For example, Malekian et al. (2020) used almost 10,000 learner-behaviour sequences before each assessment submission to predict their assessment readiness. They found that the student *sequence* of task preparation activities (e.g., downloading lectures, viewing forums, and reviewing materials) was a powerful predictor of assessment performance. However, although predictive sequences are often identified in large datasets, interpreting what these sequences mean can be difficult due to a lack of follow-up studies to validate researchers' inferences about them.

SPM can be conducted through various methods, one of which involves representing learning events as n-grams. In natural language processing, n-grams represent single or consecutive word combinations (He & Von Davier, 2016). For example, in the sentence “I take notes when I watch videos,” there are seven one-grams (e.g., “I,” “take,” “notes”), six two-grams (e.g., “I take,” “take notes”), and five three-grams (e.g., “I take notes,” “take notes when”). Applied to learning events, n-grams can represent sequences like downloading lectures, viewing forums, and reviewing materials as a three-event sequence.

SPM such as n-grams capture the sequential and temporal dimensions of SRL and offers a means to examine temporality in understanding SRL processes (Chen et al., 2018). The temporality of SRL can be analyzed by looking at factors like position, duration, frequency, and rate of events, as well as patterns of regular or irregular change over time (Molenaar & Wise, 2022). N-grams thus enable researchers to explore the order of events and can serve as a foundation for further studies on the role of temporality in performance by examining the frequency of temporal events in relation to academic outcomes.

2.4. Interpreting Sequences of Digital Events

Interpretation is the crucial next step after identifying sequences of digital events, and doing so holds both practical and theoretical importance. For example, repeatedly scrubbing forward during an instructional video (e.g., forward-forward-forward) may be found to predict learning outcomes (Kuhlmann et al., 2024), yet its meaning can vary based on the learner. Those with prior knowledge may use this behaviour to revisit or skip to key content, aligning it with their existing understanding. Others might search for a specific concept or scan through the material to ensure exposure without fully engaging with it. Without examining the underlying intent behind these actions, instructors risk misinterpreting the behaviour, attributing outcomes to the behaviour itself rather than learner characteristics. Accurate interpretation is thus necessary to derive meaningful inferences and apply insights from digital event data to better understand SRL processes in future learners.

One approach to interpret sequences involves first validating individual events before combining them in sequence. Although this method offers a structured way to analyze behaviours, it risks oversimplifying learning by ignoring how behaviours interact within the broader context. For instance, researchers have used surveys, or retrospective interviews, to validate learner goals and intentions behind specific video-watching behaviours (e.g., pausing; Liao & Wu, 2023; Seo et al., 2021). Pausing is often linked to reflection or taking a break, and forwarding indicates skimming or reorienting oneself. However, when examining a sequence of video-watching behaviours, such as pausing-forwarding, if inferences are based on the simple combination of interpretations of individual events, this might induce multiple conflicting interpretations, obscuring the nuanced interactions between events, as the meaning of each sequence can shift depending on the context.

Treating sequences of digital events as integrated parts mirrors the microgenetic method (Siegler & Crowley, 1991), which emphasizes that learning unfolds through gradual, interconnected steps. This approach highlights the importance of examining relationships between actions over time to capture the dynamic, evolving nature of learning. Just as microgenetic analysis examines learner behaviours over time, interpreting sequences requires examining overall patterns of temporal behaviours. Both approaches recognize that understanding learning requires analyzing sequences of behaviours in context, rather than isolating individual events.

3. The Current Study

In this study, we have two research questions:

RQ1: How do sequences of digital events enacted by undergraduate students with digital materials in a flipped lesson predict their achievement in instructional tasks?

RQ2: How do these predictive sequences of digital events align with student verbalizations, and what types of SRL events do they reflect?

We overcome the challenges of large-sample SPM studies and the inference validity of digital events in two ways. First, we built on n-gram methods to identify learning sequences of interest. We achieved this by capturing multiple digital events that undergraduates engaged in during a lesson from their math class and examined how these sequences predict achievement. Second, we used think-aloud protocols to obtain rich, authentic descriptions of learner thoughts, avoiding reliance on self-reports that may be affected by learner understanding and recollection. We coded these verbalizations and aligned them with sequences to create inferences that could be scaled.

These learning sequences can be observed in classrooms as potential predictors of achievement and understood through the lens of SRL. This foundation supports the development of theory-aligned algorithms as an instance of explainable artificial intelligence tools (Khosravi et al., 2022) that can be used to understand and support student learning in authentic educational settings.

4. Methods

4.1. Participants

Participants in this study were recruited from students enrolled in Precalculus Mathematics in a southeast university in the U.S. Our recruitment resulted in a convenience sample of 55 students from a single Precalculus course offered over two semesters: Fall 2020 and Spring 2021. Upon completion, each participant was compensated for their time with a 75-dollar Amazon gift card. The data from six participants were excluded from the analyses due to technical difficulties. Among the remaining 49 participants, 77.5% were female, mostly in their first and second years (87.8%), and their age ranged from 18 to 25 (Mean Age = 19, SD = 1.46). All data collection was conducted prior to the lesson taught in class each semester.

4.2. Redesigned Lesson and Event Logging Across Platforms

The research team worked with the instructors of the math department to design a lesson on ellipses. The lesson followed three phases to reflect the active learning pedagogy that the instructor adopted: pre-class (15 minutes), in-class (50 minutes), and post-class (20 minutes; see Figure 1). In the pre-class phase, learners watch a video about ellipses theory on Sakai and complete three-item homework assignments on MyMathLab, a digital learning platform provided by the textbook publisher. During the in-class phase, learners watch a lecture video on Sakai about ellipses, and solve a practice problem on Learning Catalytics, an interactive student response tool. Learners are then provided with the opportunity to review worked examples from two fictitious peers. After this optional task, all learners are given a review video where the instructor explains the steps of the practice problem. In the post-class phase, learners complete two assessment items on MyMathLab.

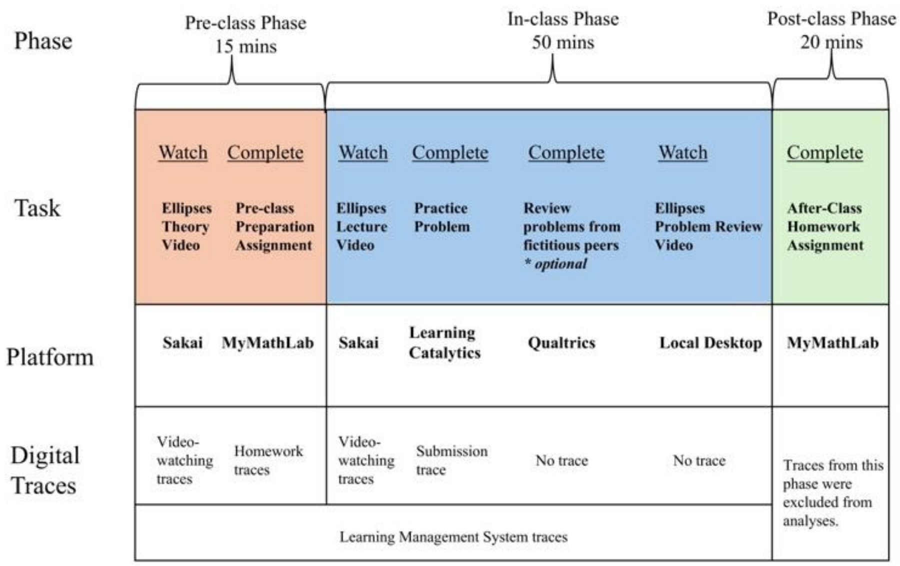

Figure 1. Study design.

During phase transitions, research assistants read the directions for the next phase and logged into the systems on behalf of learners. Within each phase, learners were instructed to engage with the tasks as they normally would during regular coursework. All platforms — including Sakai, MyMathLab, and Learning Catalytics — are part of daily instruction in this course. Learners were at liberty to organize windows across their monitors as they saw fit, with no requirements imposed by researchers, and no documentation of ways that learners chose to arrange their application windows. Digital traces were logged across these platforms, including anonymized student IDs, timestamps, and actions. We labelled activities initiated by research assistants at the beginning of learning sessions and during transitions, and chronologically linked each student's data using their anonymized ID. One important note is that the Ellipses theory and Ellipses lecture videos were hosted on the same lesson page. When a participant finishes the first video, the system automatically starts playing the second one. If the participant is not ready to move on to the next phase, they may manually pause the second video. Additionally, if a participant pauses a video for an extended period, the system may automatically resume playback. As a result of these system behaviours, there are more recorded Pause activities than Play activities, as shown in Table 1.

4.3. Study Procedure

This study happened during the global pandemic, so the entire session was conducted over Zoom. Upon entering the Zoom room, each participant was greeted and given an overview of the study. To protect learner privacy, research assistants gave participants remote-control access to their computers. After reviewing and signing the consent forms, learners were given think-aloud practice (TAP) training and opportunities to ask any questions regarding TAP. Learners proceeded to complete the lesson (with the same design as their course) on a mock Sakai site. Upon completing the lesson, participants concluded with a demographics survey and exited Zoom.

4.4. Measures

4.4.1. Think-Aloud Protocol

Participants were asked to think-aloud during the entire session. They were told to verbalize what they were thinking, feeling, and doing (Ericsson & Simon, 1993). During the learning task, if a participant was silent for more than two seconds, the researcher gave the prompt of "*Please keep talking*." The entire session was audio- and video-recorded through Zoom. The audios were professionally transcribed by *Rev.com*, and the transcripts were then coded by research assistants according to the codebook used in Greene et al. (2018). The codebook includes more than 50 micro codes nested into five macro categories, including *task codes*, *monitoring*, *domain-general strategies (DGS)*, *domain-specific strategies (DSS)*, and *assessment strategies (Assess)*. See the Appendix for code descriptions and sample texts.

Three coders participated in the coding process. After completing initial training, they began coding independently once sufficient agreement was reached. Two raters double coded all documents, and final codes were determined collaboratively between them. Given the vast number of potential codes produced in student verbalizations, percentage agreement was used to assess interrater reliability (see Bernacki et al., 2025 for details). Each pair of raters coded transcripts from a different number of participants; therefore, we calculated interrater reliability by weighting the percentage agreement across all rater pairs. The overall weighted agreement was 0.74, which is considered substantial (Hallgren, 2012).

4.4.2. Digital Traces

Student interactions with three digital platforms (i.e., Sakai, Learning Catalytics, and MyMathLab) were recorded through server log files. The log files included time-stamped records of the student interactions, such as page loads and file downloads, video-watching behaviours, such as forward and backward on Sakai; submission traces with correctness and hint usages on MyMathLab; and submission traces with correctness on Learning Catalytics. Table 1 provides a description of each trace and the total count of traces before the assessment started.

Table 1. Trace Description

	Traces	**Number of Occurrences**	**Number of People**
	Homework (HW) traces		
1	Access the first homework question	49	48
2	Check the answer of the first homework question	48	48
3	Complete the first homework question	48	48
4	Leave the first homework question	1	1
5	Access the second homework question	52	47
6	Check the answer of the second homework question	49	46
7	Complete the second homework question	49	46
8	Leave the second homework question	3	3
9	Access the third homework question	58	46
10	Check the answer of the third homework question	55	44
11	Complete the third homework question	55	44
12	Leave the third homework question	3	3
	Video-Watching traces		
13	Load video on Sakai	184	47
14	Play (a video on Sakai)	622	48
15	Reengage (e.g., accesses a video after 5 minutes of inactivity)	8	7
16	Restart (e.g., moves back to the beginning of the video)	12	10

17	Rewind	109	21
18	Complete (watching a video)	65	42
19	Engage (e.g., accesses a video for the very first time)	94	48
20	Forward	379	41
21	Pause	505	43
	Learning Management System (LMS) traces		
22	Read lesson page	108	45
23	Read syllabus	2	2
24	Navigate links (e.g., Pearson, Piazza, or Math Centre on Sakai)	4	3
	Learning Catalytics (LC) traces		
25	Submit LC question	47	38
	Total	2,611	

4.4.3. Assessment Measure

In the post-class phase, students answered two quiz questions, the first question (four steps) and the second question (two steps). Students were allowed to attempt each question up to three times. We used item response theory (Bock & Gibbons, 2021) to produce math performance scores for each student. Each question step was coded using the observed attempt count (OAC) approach for the graded response model (Bergner et al., 2019). Using this coding, a correct answer on the first attempt receives the highest score (in our case, 3); a correct answer on the second attempt receives the second-highest score (2), and the pattern continues, with an incorrect answer receiving 0. Negative numbers such as "–2" means that the student made two attempts, and they got it wrong in the second attempt. The most common response pattern was 3.3.3.3.0.0 (N = 13), meaning correct answers on the first attempt of the first four steps, and incorrect answers for the last two steps. Given the relatively small sample, we estimated a simplified graded response model that treated all question steps as equally reliable indicators of math performance. This is because our sample size was not sufficient to obtain stable estimates of the GRM discrimination parameters for each item, with some items showing divergent discrimination parameters (> 10) and standard errors. To address this issue, we constrained the model so that all items had equal discrimination parameters, resulting in a so-called tau-equivalent model. To test whether the tau-equivalent model was suitable for our data, we compared it to the unconstrained GRM. The likelihood ratio test was significant ($\chi2(5) = 17.94$, $p = 0.003$) but the BIC statistic indicated that the tau-equivalent model represented a better compromise between fit and parsimony. The EAP from the two models were correlated at $r = 0.96$. Ultimately, we selected the tau-equivalent model because the sample size was not sufficient to obtain stable estimates of the GRM discrimination parameters, with some items showing divergent discrimination parameters (> 10) and standard errors.

Math performance scores were produced using the expected *a posteriori* method (Bock & Gibbons, 2021). In this sample, student scores ranged from –1.9 to 1.56 (M = 0, SD = 0.89) with an overall reliability of 0.77. The marginal reliability for the math achievement scores was computed using the conditional reliability coefficient presented in Nicewander (2018; Equation 30), averaged over the range of the latent trait. For implementation, we used the "marginal_rxx" function of the "mirt" software package (Chalmers, 2012).

4.5. Analytical Strategies for RQ1: Identifying Sequences Most Predictive of Task Performance

Our overall analytical approach to RQ1 was as follows. First, we used text-based approaches from natural language processing (Dong & Pei, 2007) to encode digital trace sequences observed during the pre-class and in-class learning phases of the lesson. Second, we used LASSO regression (Franklin, 2005) to identify the sequences most predictive of student math performance during the post-class assessment phase of the lesson. This same overall approach has been used in similar contexts to predict student performance based on their in-task behaviours (He & Von Davier, 2016; Ulitzsch et al., 2022, 2023). We describe these two steps in detail below.

4.5.1 Step 1: Data Preparation

To extract sequential information for the digital trace data, we identified all sequences (*n-grams*) of varying lengths observed across all student event logs. This was implemented using the stringr package in R (Wickham, 2023). A major decision in our study was determining the length of the sequences (e.g., how many events to include in a sequence). We set the maximum length at *three,* based on the following considerations.

First, we examined the sparsity (the proportion of zeros in the sequence counts across students). The sparsity reached 90.1% when we extended the sequences to five events. Based on this, we could extract sequences with a maximum length of *five* events. Second, we aimed to capture meaningful and interpretable sequences. Longer sequences often create nested sequences, reducing interpretability by introducing redundancy. For example, all five-event sequences are composed of four-event sequences plus one additional event, and four-event sequences are similarly built from three-event patterns. To illustrate, the three-event sequences “pause-backward-pause” and “backward-pause-backward” are both nested within the four-event sequence “pause-backward-pause-backward.” If we find that the four-event sequence is positively related to performance, it becomes difficult to determine whether the predictive value lies in the individual three-event sequences or in the overall four-event sequence. Third, longer sequences also represent extended periods of time, complicating interpretation. For instance, the average duration is 90.03 seconds for three-event sequences, 127.57 seconds for four-event sequences, and 164.11 seconds for five-event sequences. These longer durations may overlap with multiple verbalizations, making it difficult to determine whether individual self-regulated learning processes or combinations best explain the sequence.

Given these considerations, sequences of up to *thre*e events offer a balance — rich enough to capture meaningful patterns without repetitive observations of nested patterns or making it overly complex or difficult to interpret. After removing traces related to research assistant activities, 388 unique participant-generated events or event sequences of up to three events remained.

4.5.2. Step 2: Variable Selection and Hierarchical Linear Regression

The process of selecting sequence-based predictors of math performance involved a combination of automated variable selection through LASSO regression (implemented in glmnet package in R; Friedman et al., 2010) followed by manual refinement based on effect sizes (zero-order correlations). In general, variable selection methods such as the LASSO can be highly automated when sample sizes are sufficiently large (e.g., using cross-validation to compare different penalty functions and their tuning parameters). Due to the small sample size in the present study, we adopted a more “bespoke” approach to variable selection.

First, we excluded sequences that were either too sparse (performed by only one or two participants) or too dense (performed by all but one or two participants) to ensure that the sequence indicators did not act as proxies for specific individuals. As a result, we removed 237 sparse events or event sequences and six dense events or event sequences from the initial set of events and sequences of up to three events, leaving a total of 146 events or event sequences.

Additionally, as variable selection with the LASSO is well-known to be affected by multicollinearity among the predictors (e.g., Freijeiro-González et al. 2022), we mitigated this issue by omitting collinear (r = 1) event sequences during data preparation. Specifically, we identified cases where longer sequences had the exact same support — defined as the number of participants who performed the sequence — as shorter ones. For example, the sequences *accessHw1_checkAnswerofHw1* and *accessHw1_checkAnswerofHw1_ completeHw1* had identical support values. This can occur when participants transition rapidly from checking the answer to completing the homework, causing both the shorter and longer sequences to appear equally often. These longer sequences were excluded because they did not offer additional insight into participant behaviour and could introduce confounding effects in the subsequent variable selection process. In total, we removed 28 such longer sequences, reducing the set from 146 to 118 — the final set of event and event sequence variables used in our analysis.

Then, the penalty of the LASSO regression was relaxed until the first 20 predictors entered the model. To address this issue of possible multicollinearity further, we also conducted a sensitivity analysis using the elastic-net, which is robust to multicollinearity (Zou & Hastie, 2005). Specifically, we compared the results of the LASSO (elastic-net with a tuning parameter of alpha = 1) to the results obtained using α values from 0.5 to 1.0. As expected, different α values resulted in slightly different sets of selected predictors, particularly given the small sample size and potential multicollinearity among variables. These differences reflect how the model penalizes collinear or weakly predictive variables under different constraints. However, after applying a post-hoc pruning step — removing any predictor with a zero-order correlation of less than 0.2 with math performance — the final sets of predictors were included across the tested α values. This consistency suggests that our key predictors are stable and meaningfully related to performance, even under varying model constraints.

Last, we pruned this set of predictors by omitting any sequence whose zero-order with math performance had an absolute value of less than 0.2. We used the zero-order correlations as a post-hoc way of pruning the LASSO-selected sequences based on their practical significance rather than statistical significance, because of our focus on understanding the practical significance of the sequences. The rationale here was that we wanted our final set of sequence-based predictors to each be meaningfully related to math performance on their own, as we intended to interpret each sequence individually. This practical measure communicates the *magnitude* of the relationship, not just whether a significant relationship exists. In educational contexts where sample sizes are modest, prioritizing effect size and confidence intervals can provide more informative guidance than significance tests (Kirk, 1996; Kline, 2013; Kraft, 2020; Maher et al., 2013). Our choice of r = 0.2 as a cut-off for practical significance is comparable to the overall correlation between hours of homework and math/science achievement reported in the meta-analysis of Fan et al. (2017). It also aligns with Ferguson’s (2016) recommendation that r = 0.2 represents a minimum effect size worth interpreting in applied research contexts such as education.

4.6. Analytical Strategies for RQ2: Interpreting Predictive Sequences with Verbalizations

To answer RQ2, we used two alignment methods, as our data included both single events and event sequences (i.e., two-event and three-event sequences). Single events are essentially instantaneous time points with no duration. To align them with verbalizations, we identified verbalizations that occurred within a reasonable temporal proximity. Our previous work (Bernacki et al., 2025) showed that by using a 20-second window, most events could be matched with verbalizations, and further analyses indicated that most of these events aligned with a dominant verbalization. Similar approaches have been used in other studies (e.g., Baker et al., 2020). Therefore, in addition to direct overlap — where the digital event occurs within the time span of a verbalization (i.e., Type 1 verbalizations in Figure 2) — we also considered verbalizations that ended within 10 seconds before the digital event (Type 2 verbalizations in Figure 2) and those that began within 10 seconds after the digital event (Type 3 verbalizations in Figure 2) as aligned.

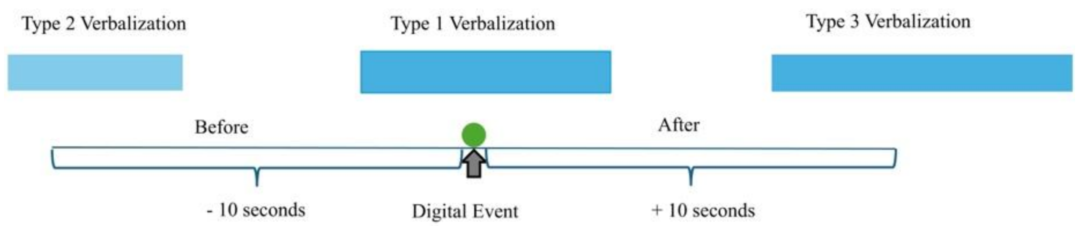

Figure 2. Alignment method for single events.

For event sequences, both the sequences and the verbalizations span a duration. We identified six types of temporal relationships between verbalizations and sequences (Figure 2). For example, for a three-event sequence (see the white bar with three green dots in Figure 3):

Type 1A: The verbalization starts and ends within the sequence's duration.

Type 1B: The verbalization starts before and ends after the sequence's duration.

Type 2: The verbalization starts before and ends within the sequence.

Type 3: The verbalization starts within and ends after the sequence.

Type 4: The verbalization ends before the sequence begins.

Type 5: The verbalization starts after the sequence ends.

In our analysis, Types 1A and 1B were considered fully aligned, whereas Types 2 and 3 were considered partially aligned due to their partial overlap with the sequence. For instance, Type 2 overlaps with the beginning of the sequence, and Type 3 with the end. We included Types 1–3 in our subsequent analyses and excluded Types 4 and 5 due to their lack of temporal alignment. Including Type 4 and 5 verbalizations would also require setting a cut-off distance to infer relevance — a methodological challenge not yet adequately addressed in the current literature.

Figure 3. Several types of verbalizations in relation to the timing of a sequence.

Note: Type 1A, 1B, 2, and 3 verbalizations are all considered aligned to event sequences. Numbers 1, 2, 3 denotes different events in the three-event sequence.

Next, to interpret the predictive sequences through verbalizations, we tabulated the distribution of verbalizations (including Types 1A, 1B, 2, and 3) aligned with the sequence and compared it to the SRL macro code distribution in the learning session, using chi-square tests of proportions. Since these tests require at least 80% of the expected frequencies in the contingency table to be greater than 5 (McHugh, 2013), one of the six predictive sequences did not meet the requirements. Therefore, we also conducted Fisher's exact test, which is appropriate to analyze contingency tables with sparse or unbalanced distributions because – unlike the chi-square test that relies on approximations requiring larger sample sizes – Fisher's exact test employs an exact procedure using hypergeometric distribution. It can thus theoretically be applied to larger contingency tables and remains valid for all sample sizes, offering a more reliable alternative to chi-square tests when analyzing sparse data distributions (Kim, 2017).

Finally, after chi-square tests of proportions and Fisher's exact test, we examined the student qualitative data (i.e., exact verbalizations) to gain deeper insights into SRL processing. For each predictive sequence, we presented modal cases of reflecting common SRL macro processes, supplementing with the exact quotes and micro codes to add further nuance.

5. Results

5.1. Descriptive Statistics

There were 118 events or event sequences after cleaning the sparse and dense n-grams, including 14 single events, 40 two-event sequences, and 64 three-event sequences. The frequency of events or event sequences ranked from three to 505. The number of people who generated those sequences range from three to 47. See Table 2 for the top 10 frequence events or event sequences.

Table 2. Top 10 Frequence Events or Event Sequences

		Number of Occurrences	Number of People
1	pause	505	43
2	play-pause	435	43
3	forward	379	41
4	forward-play	344	40
5	pause-forward	317	40
6	pause-forward-play	294	39
7	play-pause-forward	289	40
8	forward-play-pause	241	34
9	Load a video	184	47
10	Rewind	109	21

5.2. RQ1: Identifying Sequences Most Predictive of Task Performance

See Table 3 for the event or event sequences that have an absolute correlation of at least 0.2 with the Math performance. These six selected events or event sequences were included using all three levels of alphas in our sensitivity analyses. The six predictors explained a substantial portion of the variance in later task performance (R^2 = 0.371). The six most predictive sequences included three predictors with a negative correlation with performance:

1) Rewind (r = –0.291)
2) Pause, rewind, pause (r = –0.333)
3) Access the third homework question and leave it ("Access HW3 and Leave"; r = –0.376)

The *three* sequences positively correlated with performance were the following:

4) Read lesson page and read another lesson page (r = 0.275)
5) Access the second homework question and check the answer ("Access HW2 and Check"; r = 0.285)
6) Complete the second homework question, access the third homework question, and check the answer of the third homework question ("Complete HW2, Access HW3, and Check HW3"; r = 0.369)

All VIF values were below 2.5, indicating that multicollinearity was not a concern. See Figure 4 for a summary plot of the six selected events or event sequences across all participants.

ISSN 1929-7750 (online).

Table 3. Six Selected Events or Event Sequences in the Regression Model

	Event or Event sequences	Correlation Coefficients	95% CI	Standardized Regression Coefficients	95% CI	P-values in the Regression Model	VIF
1	Rewind	–0.291	[–0.529, –0.011]	–0.059	[–0.422, 0.304]	0.104	2.161
2	Pause, rewind, pause	–0.333	[–0.562, –0.058]	–0.307	[–0.672, 0.058]	0.097	2.185
3	Access HW3 and Leave	–0.376	[–0.595, –0.106]	–0.376	[0.648, –0.103]	0.008	1.220
4	Read lesson page and read another lesson page	0.275	[–0.007, 0.516]	0.083	[–0.185, 0.353]	0.533	1.189
5	Access HW2 and Check	0.285	[0.004, 0.524]	0.253	[–0.049, 0.554]	0.099	1.494
6	Complete HW2, Access HW3, and Check HW3	0.369	[0.098, 0.589]	0.041	[–0.272, 0.354]	0.793	1.605

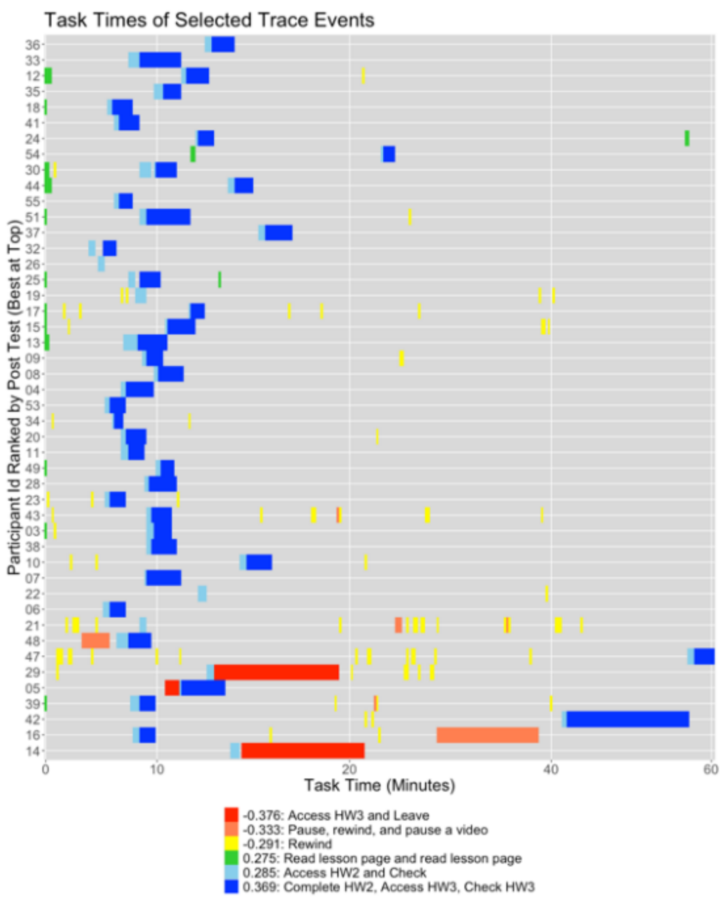

Figure 4. Summary plot of six selected events or event sequences across participants.

Note: Task time was calculated as the time of the last event minus the time of the first event. The participant who took the longest to complete the task did so in approximately 59 minutes, which is why we used 60 minutes as the endpoint.

5.3. RQ2: Interpreting Predictive Sequences With Verbalizations

We next examined how events or event sequences that predicted later task performance would align with verbalization data. Thereafter, we examined how SRL codings and qualitatively appraised raw data might indicate how these digital events reflect SRL processes. See Table 4 for the total number of six selected events or event sequence occurrences among all participants, the number of participants engaging in each predictive event or event sequence, the average duration of predictive event or sequence, and total number of Type 1 verbalizations (full alignment) and Type 2 and 3 verbalizations (partial alignment) for each sequence. We then summed these verbalizations for each sequence and examined the distribution across five different SRL macro processes (see Table 5).

Table 4. Description of Six Events or Event Sequences

	Event or Event Sequence	# of Occurrences	# of People	Average duration (seconds)	# of Type 1	# of Type 2 and 3	Total of Type 1–3
1	Rewind	109	21	NA	29	21	50
2	Pause, rewind, pause	6	5	50.667	10	1	11
3	Access HW3 and Leave	3	3	64.228	11	1	12
4	Read lesson page and read another lesson page	15	14	6.0667	2	0	2
5	Access HW2 and Check	48	46	17.529	58	8	66
6	Complete HW2, Access HW3, and Check HW3	40	40	70.451	82	17	99

Note: For event sequences, Type 1 verbalizations include both Type 1A and Type 1B verbalizations in Figure 3. Type 1B verbalizations were rare — among the five-event sequences, only one instance of a Type 1B verbalization was found (for the sequence Access HW2 and Check), and none were observed for the other sequences. This is likely because verbalizations rarely begin before and end after an entire sequence.

We hypothesized that the verbalization distribution with each event or event sequence should be the same as the verbalization distribution within the overall learning session. Significant differences between observed and expected distributions would suggest that a sequence is associated with the frequent co-occurring SRL processes reflected in the verbalization. Chi-square tests of proportions revealed that the distribution of the SRL process for one sequence was significantly different from the overall distribution: *Access HW2 and Check* (r = 0.285; Table 5). The SRL macro processes distribution for four sequences did not significantly differ from the SRL macro processes in the overall learning session. Due to small sample sizes, chi-square tests of proportions were not feasible for one sequence: *Read lesson page and read another lesson page*.

Table 5. Distribution of Type 1, 2, 3 Verbalizations Across the Five SRL Macro Processes

	Event or Event Sequence	Task Codes	Monitoring	DGS	DSS	Assess	Total	Chi-square tests
1	Rewind	3	11	31	5	0	50	$p = 1$
2	Pause, rewind, pause	1	5	3	2	0	11	$p = 1$
3	Access HW3 and Leave	1	4	2	5	0	12	$p = 1$
4	Read lesson page and read another lesson page	1	1	0	0	0	2	Cannot test due to small sample size
5	Access HW2 and Check	0	10	12	40	4	66	$\chi^2 = 37.63$, $p = 0.0005$
6	Complete HW2, Access HW3, and Check HW3:	0	19	23	51	6	99	$p = 1$
	Total verbalizations of the six sequences	6	50	71	103	10	238	
	Total verbalizations of the learning session	103	908	621	687	121	2,440	

Note: We only counted verbalizations coded as one of the five macro processes in the codebook. The chi-square test results reflect tests of proportions, comparing the distribution of the five macro processes within each event or sequence to the overall distribution of macro processes across the learning session (the last row). As some expected cell frequencies were very small, we estimated the p-value of the chi-square test using a Monte Carlo simulation with 2,000 replicates to ensure statistical rigour.

Next, we conducted Fisher's exact tests for events or event sequences (Figure 5). The Residuals Heatmap calculates standardized residuals, showing the strength and direction of deviations from expected frequencies under independence. Next, we drew on qualitative data to further interpret the findings, focusing on *modal codes* — those significantly associated with each event or event sequence-based on the chi-square tests or Fisher's exact tests. These modal codes, along with representative participant quotes, are presented in Table 6.

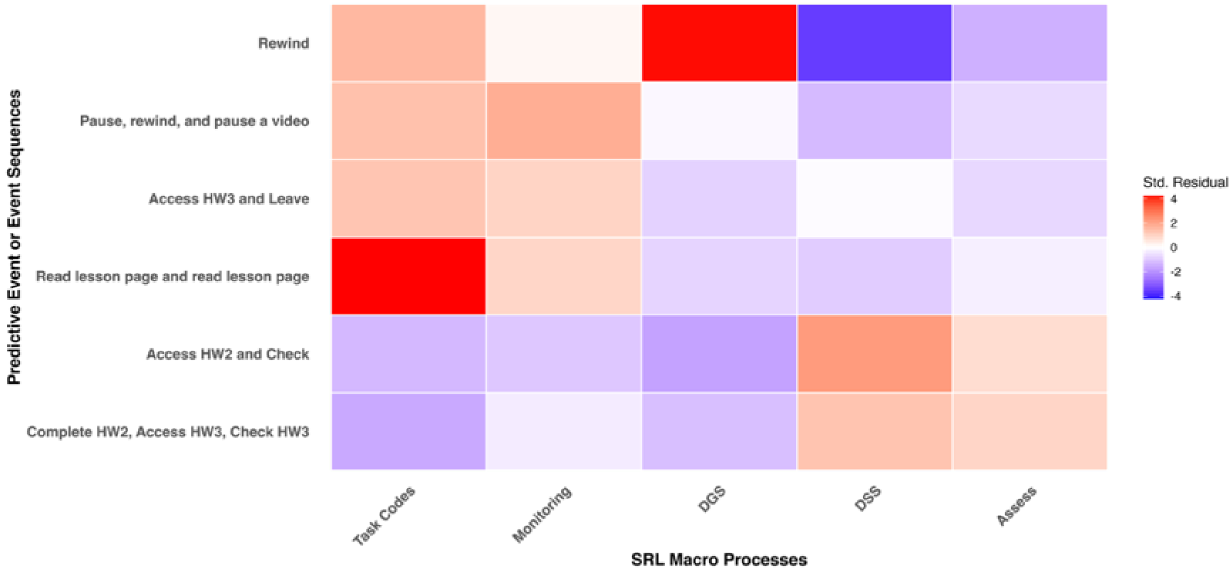

Figure 5. Fisher's exact test residuals heatmap.

Note: Purple indicates negative residuals (fewer observations than expected). Red indicates positive residuals (more observations than expected). White indicates values close to expected. The intensity of the colour represents the magnitude of the deviation.

- Rewind

The chi-square test showed that the SRL macro processes distribution for an event did not significantly differ from the SRL macro processes distribution in the overall learning session. Fisher's exact test indicated that the event was more associated with DGS (strong) and Task Codes (moderate) than expected and less associated with DSS (strong) and Assess (moderate) than expected. As an example of the modal codes, Participant 47 enacted Rewind 6 times within 5 minutes, during which they verbalized DGS (e.g., rereading) for three times and Task Codes (e.g., Sub-goal) once. Participant 12 used DGS (e.g., Rereading) for the only time they engaged in Rewind.

- Pause, Rewind, Pause

The chi-square test showed that the SRL macro processes distribution for this sequence did not significantly differ from that of the overall learning session. Fisher's exact test indicated the event sequence was more associated with Task Codes (moderate) and Monitoring (moderate) than expected and less associated with DSS (moderate) and Assess (moderate) than expected. As an example of the modal codes, Participant 16 enacted this sequence one time, during which they mainly verbalized Monitoring strategies (i.e., CE-, JOU+, SKA, JOU-, and progress monitoring) five times.

- Access HW3 and Leave

The chi-square test showed that the SRL macro processes distribution for this sequence did not significantly differ from that of the overall learning session. Fisher's exact test revealed the event sequence was more associated with Task Codes (moderate) and Monitoring (moderate) than expected, and less associated with DGS (moderate) and Assess (moderate) than expected. As an example of the modal codes, participant 14 realized they were running out of time when they enacted this sequence and verbalized Monitoring (e.g., Task Difficulty). Similarly, participant 29 engaged in Monitoring two times (i.e., negative Judgment of Understanding and negative Feeling of Recognition) when they enacted this sequence.

- Read Lesson Page and Read Another Lesson Page

A chi-square test could not be performed due to the small sample size and unequal distributions in the columns. Fisher's exact test result shows that the event sequence was more associated with Task Codes (strong) and Monitoring (moderate) than expected, and less associated with DGS and DSS (moderate) than expected (Figure 5). As an example of the modal codes, participant 44 used Task Codes (e.g., Sub-goal) once and Monitoring (e.g., negative Feeling of Knowing) once.

- Access HW2 and Check

The chi-square test showed that the SRL macro processes distribution for this sequence differed significantly from that of the overall learning session. When learners enacted this sequence, they often engaged in DSS (Table 6). Fisher's exact test result shows that the event sequence was more associated with DSS (moderate) and Assess (weak) than expected, and less associated with Task Codes, Monitoring, and DGS Assessment (moderate) than expected (Figure 5). As an example of the modal codes, participants 25 and 54 only used DSS (e.g., MPS) when they enacted this sequence.

- Complete HW2, Access HW3, and Check HW3

The chi-square test showed that the SRL macro processes distribution for this sequence did not significantly differ from that of the overall learning session. Fisher's exact test result shows that the event sequence was more associated with DSS (moderate) and Assessment strategies (weak) than expected, and less associated with Task Codes (moderate), Monitoring (moderate), and DGS (moderate) than expected (Figure 5). As an example of the modal codes, participants 3 and 4 only engaged in DSS (e.g., MPS) when they enacted this sequence.

Table 6. Verbalizations of SRL Macroprocesses in Six Selected Events or Event Sequences

Digital Event Sequence	Verbalizations	SRL Macro and Micro Processes
1. Rewind	P12: I'm going to go a little bit back (and start watching)	DGS: RR
	P47: I'm going to rewind to figure out why I'm doing this. I missed... What we're trying to solve at this	Task Codes: Sub_goal
	P47: And then I'm going to factor out the 16. So that'll be 16 Y squared minus the square	DGS: RR
	P47: 6 over 2 squared, which is 3 squared	DGS: RR
	P47: So I have 6 over 2 squared, which [rewind 00:36:26] squared. So for the X terms, I'll have 6 over 2	DGS: RR
2. Pause, Rewind, Pause	P16: so the video kind of confused me	Monitoring: CE–
	P16: for the most part I understood the concept of it	Monitoring: JOU+
	P16: that's where I struggle most on is completing the square	Monitoring: SKA
	Reading notes	
	Solving math problems	
	P16: I don't know, I'm a little confused by this part	Monitoring: JOU–
	Reading notes	
	Solving math problems	
	Setting a goal to watch video	
	P16: How many learning catalytics questions are there, just one okay	Monitoring: MP
3. Access HW3 and Leave	P14: this one's going to be a little trickier	Monitoring: TD
	Reading notes	
	Solving math problems	
	P29: It's a little bit confusing	Monitoring: JOU–
	P29: I'm not very familiar with a lot of these	Monitoring: FOR–
	Reading notes	
	Solving math problems	
4. Read Lesson Page and Read Another Lesson Page	P44: I want to watch the video first	Task Codes: Sub_goal
	P44: because I don't remember anything about ellipsis from high school math	Monitoring: FOK–
5. Access HW2 and Check	P25: the foci lie on a line called then, I believe, it was the minor axis	DSS: MPS
	P54: the major axis	DSS: MPS
6. Complete HW2, Access HW3, and Check HW3	P3: it's the bigger value. So that's 25 and then it's squared, so that would just be five. And then B is 3 cause three squared is 9	DSS: MPS
	P4: I'm going to rewrite that on my notepad. X squared over 25 plus Y squared over nine equals one. The value of A is…	DSS: MPS

Note: Exact quotes in italics are omitted to reduce word count.

6. Discussion

In this study, we employed sequence mining and variable selection techniques to 1) identify sequences of digital traces predictive of performance on a college algebra task, and 2) apply self-regulated learning as a lens and verbalizations of SRL processes as a resource to understand how sequences of behaviour reflect productive, math-specific SRL processes. The sequences selected explained substantial variance in task achievement and provide some evidence of how sequenced digital event data reflect self-regulated learning processes.

6.1. Using Digital Sequences of Events to Predict Performance

Primarily, our study underscores the importance of complexity in SRL frameworks, inclusive of temporally sequential processes (e.g., Molenaar & Järvelä, 2014, drawing on Winne and Hadwin, 1998). Most predictive patterns (five out of six) were identified as event sequences rather than single events. This finding underscores the value of examining the sequential and temporal dimensions of SRL processes, as they provide deeper insights into how students navigate complex learning tasks over time, and further examining them in context, under observable task conditions (Winne & Hadwin, 1998).

We identified two video-watching behaviour sequences related to performance outcomes. The sequences *Rewind* and *Pause, rewind, pause* were negatively predictive of performance. These findings largely align with existing literature on learning designs in STEM domains. Chi and Wylie (2014) conceptualized video-watching behaviours as forms of active engagement in their Interactive, Constructive, Active, and Passive (ICAP) framework. Seo et al. (2021) further noted that video-watching behaviours such as forwarding, pausing, and rewinding often reflect distinct intentions tied to engagement goals. Rewinding, for instance, may indicate attempts to recall or clarify content due to distraction or initial misunderstanding, which could explain why sequences involving pausing and rewinding predicted poorer performance. Another negative predictor, *Access HW3 and leave it*, indicates task abandonment, a behaviour often associated with lower performance. Kasakowskij et al. (2023) similarly found that task abandonment behaviours were predictive of disengagement from the course, evidenced by the absence of further activity in the learning management system.

In contrast, two sequences were positively predictive of performance and reflect task engagement: *Access HW2 and Check* and *Complete HW2, Access HW3, and Check HW3*. Both sequences reflect task engagement. Given that the homework included three questions, progressing to the second and third questions indicates active engagement sufficient to make timely progress through a task, and this is positively related to performance. The other positive predictor, *Read lesson page and read another lesson page*, was described by students as reflective of a repeated task definition process (i.e., Bernacki et al., 2025, with the process deriving from Winne and Hadwin, 1998). This process was previously found to be positively related to task performance (Greene et al., 2012), and observation of the same positive relationship provides additional evidence that engagement in task definition is associated with successful performance on academic tasks, here including mathematical problem-solving in flipped classroom settings.

6.2. Interpreting Predictive Sequences in Context

In order to align digital traces to verbalizations known to reflect SRL processes and examine what such digital events or event sequences might reflect, we then quantified coded verbalizations and used chi-square and Fisher's exact tests to determine if the distribution of TAPs across SRL macroprocesses differed from the distribution in the overall learning session. Among the six sequences that explained variance in task performance, the sequence *Access HW2 and Check* consistently reflected cognitive strategy use (i.e., domain-specific strategies at a macroprocess level, and predominantly mathematical problem-solving at the microprocess level). These SRL processes reported by undergraduate learners align closely with well-established metacognitive and cognitive processes in SRL theory; for example, strategy use is frequently observed to explain variance in performance on academic tasks involving science (Azevedo et al., 2004), and here, mathematics.

We further corroborated and enriched these findings with some modal cases from qualitative analysis of learner verbalizations. The modal case showed that rewinding tends to co-occur when someone is enacting a general learning strategy, often in response to a metacognitive judgment based on task features or based on the product of an assessment strategy. We also leveraged video data as an additional context when needed. For instance, to understand why learners abandoned the third homework question, verbalizations alone were insufficient, and the video data provided insights into the timing and rationale behind their actions. These multimodal data provided valuable context, allowing us to interpret sequences in situ and supporting contextual SRL assumptions (Kuhlmann et al., 2024; Wiedbusch et al., 2023; Winne & Hadwin, 1998).

Multimodal data such as video helped contextualize learner decisions in some cases; however, other interpretations required examining how specific behavioural patterns unfolded alongside nearby digital events and verbalizations. Understanding the meaning behind such patterns often depends on situating them within the surrounding sequence of digital events and supplementing them with richer data sources. For example, the sequence of pausing, rewinding, and pausing again frequently co-occurred with language reflecting monitoring and attention to task details. These behaviours appear to indicate student efforts to monitor their knowledge products in relation to task goals, aligning with the "comparing against standards" component of Winne and Hadwin's (1998) COPES model. Still, these interpretations remain tentative and must be corroborated with more robust evidence, such as verbal transcript data. Similarly, the co-occurrence of lesson page reading with task-related codes and monitoring suggests that students consider task goals and adjust their behaviours accordingly — even when there is no visible product of their engagement. It is an ongoing challenge to capture less observable internal processes like metacognitive judgments that precede corresponding metacognitive control strategies they are theorized to warrant when data are constrained to digital events that log interaction with resources that largely reflect strategy use (Bernacki et al., 2025). To better understand such processes, it is necessary to examine transcript data and possibly preceding events, tracing backward until a knowledge product or an evaluative judgment becomes apparent. Mixed methods are essential for this type of contextualized analysis.

6.3. Practical Implications

Together, these findings suggest that predictive digital sequences — when interpreted through the lens of SRL theory and contextualized with learner verbalizations — can inform instructional design and early intervention. Behavioural indicators such as repeated pausing, rewinding, and task abandonment could signal learning difficulties or disengagement for some learners, and documenting them could draw the attention of instructors, or even automated check-ins to offer support to the learner. These behaviours offer valuable insights into both the demands of the learning task and the learner's experience and opportunities to support learning, if the issues learners encounter during traced moments can be understood.

Researchers who work with instructors can analyze pause and rewind sequences alongside corresponding learner verbalizations to identify concepts that students frequently revisit. This can reveal areas of confusion and inform targeted adjustments to instruction or practice. These insights are especially useful in flipped classroom models (Weinberg & Thomas, 2018), where content is delivered before class time, and noticing struggle during asynchronous tasks can inform how time in synchronous sessions is best dedicated to areas that gave students pause. Instructors may use such data to inform whole group instruction when pause points are common to all, and thereafter also use behavioural traces to identify students who could benefit from additional support on specific topics and provide it through small-group instruction during class.

These findings also have implications for the design of learning systems. Instructional designers can set triggers based on a certain number of rewinds or pauses to prompt learners to report on their challenges and offer them alternative resources if their pauses reflect struggle. Including SRL prompts such as, "What are you trying to understand by rewinding?" can transform passive behaviours into moments of reflective engagement and self-regulation (e.g., Moos & Bonde, 2016). Additionally, patterns of homework disengagement — such as leaving assignments without returning — can serve as early warning signs (Gurung et al., 2025). Learning systems could flag these behaviours and instructors can then personalize recommendations to nudge students to attend office hours or explore additional course resources.

Importantly, we do not suggest discouraging behaviours like pausing or rewinding. Instead, we recommend interpreting these behaviours as potentially meaningful signals that can guide timely and supportive intervention. Our findings underscore the need for learning environments that support contextualized, theory-informed analytics and provide both instructors and students with actionable feedback to foster metacognitive awareness and productive engagement.

7. Limitations and Future Directions

In this study, we used LASSO regression for variable selection due to the substantial number of potential predictors and our relatively small sample size. This approach reduced model complexity and allowed us to focus on a subset of relevant predictors. However, because LASSO is a data-driven selection method, we did not follow-up with traditional significance tests on the selected predictors, as this would raise issues of post-selection inference; that is, we had already excluded predictors with smaller correlations. Alternatively, if we had tested all predictors without LASSO, we would have needed to correct for multiple comparisons (e.g., using false discovery rate control) across more than 100 tests, which would have greatly reduced statistical power given our limited sample size. This trade-off highlights the inherent tension between variable selection and valid statistical inference, and our findings should be interpreted with this limitation in mind.

One limitation of our study lies in the challenges of data collection. Many traces captured in this study — such as tracking question access or submission — serve as indicators of task engagement, but they provide limited insight into learners' active cognitive or metacognitive engagement during the learning process. Although task engagement reflects objectively observable behaviours (e.g., whether learners interact with course materials), it does not necessarily reveal the deeper processes such observable events reflect, such as self-regulation. A potential direction for future research is the development of educational technology tools that 1) are genuinely useful for learners and 2) produce digital traces that can be inferred to validly capture deep self-regulatory processing.

Another limitation concerns our approach to data analysis. Effective sequence interpretation requires aligning digital and verbal data. In this study, verbalizations were considered aligned with a sequence if they occurred within or overlapped the sequence's start and end times. However, verbalizations in reasonable proximity (e.g., Types 4 and 5 TAPs in Figure 5) may still be relevant. What constitutes a “reasonable distance” can abide multiple definitions. We learn much from constraining our observation when we calculate the percentage of sequences matched with verbalizations and focus on digital events with substantial representation on a second channel of data (Bernacki et al., 2025), but researchers will continue to need to test the implications of different definitions on what can be inferred from the same data (Fan et al., 2023).

In addition, we examined only the occurrence of sequences, without considering their temporality (duration, position within an extended task), and how these relate to later learning or performance. Prior research indicates that the timing of accessing course material (e.g., before, during, on the day of, or after a lesson) affects learning outcomes (Plumley et al., 2024). A similar effect may apply to the timing of learning events within shorter tasks (Azevedo, 2007). Equally important is the duration of sequences, which refers to the length of time students engage in a particular sequence of behaviours. Nystrand et al. (2003) found that students in high-track classes engaged in lengthier in-depth discussions than those in lower-track classes. Thus, it is vital to examine whether longer engagement in similar sequences is associated with improved performance outcomes across students and learning contexts.

8. Conclusion

Analyzing longer and more complex combinations of SRL events is necessary to study the contingent, sequential, and contextual processes described in SRL models (Winne & Hadwin, 1998). Digital traces and sequences offer the potential to capture theses complex SRL events, and foundational research involving multimodal data need to be conducted at scale to include sufficient data to validate inferences drawn from such data and test complex assumptions in SRL frameworks (Winne, 2017; 2020).

Our study demonstrates that complex SRL processes are observable in authentic tasks sampled from naturalistic higher education coursework. These processes related to learning outcomes, corroborating prior research documenting associations between SRL and performance (e.g., Dent & Koenka, 2016). It also demonstrates that student verbalizations are a useful data channel to understand complex self-regulated learning processes. This aligns with Winne’s (2020) call for validated digital traces that both support inference-making and allow researchers to examine complex SRL processes at scale. Our study further highlights the importance of interpreting SRL processes within context — leveraging multimodal data and attending to adjacent digital events. Future researchers should focus on capturing actionable sequences and analyzing their temporal and contextual dimensions to generate valid and meaningful insights. Only with a robust understanding can we progress to the next phase in learning analytics: effective intervention (Winne, 2017, 2022).

Achieving this goal requires collaboration among researchers, textbook developers, and instructors. Researchers provide data-driven insights, developers create platforms to capture malleable learning traces, and instructors align these tools with pedagogical needs. Such partnerships will enable the tracking of specific, actionable learning traces and help identify targeted strategies to enhance student engagement and outcomes.

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 and Acknowledgements

This research was supported by U.S. National Science Foundation Award DRL 1920756. The opinions, findings, and conclusions, or recommendations expressed are those of the authors and do not necessarily reflect the views of the U.S. National Science Foundation. The author team wishes to acknowledge the Math Education team (led by Miranda Thomas) who handled the design of the focal course and procurement of and reporting on lesson materials. Additional acknowledgement goes to our collaborators in the University Registrar, Learning Centre, and Information Technology Services — Learning Management System offices who provided data, and co-principal investigators Kathleen M. Gates and Abigail T. Panter, who contributed to the conceptualization of the proposals that funded this work. The authors have no competing interests to declare. Data, study materials, and code for analysis are available at https://osf.io/5rtq9/.

References

Azevedo, R. (2007). Understanding the complex nature of self-regulatory processes in learning with computer-based learning environments: An introduction. *Metacognition and Learning*, *2*, 57–65. https://doi.org/10.1007/s11409-007-9018-5

Azevedo, R., Guthrie, J. T., & Seibert, D. (2004). The role of self-regulated learning in fostering students’ conceptual understanding of complex systems with hypermedia. *Journal of Educational Computing Research*, *30*(1–2), 87–111. https://doi.org/10.2190/DVWX-GM1T-6THQ-5WC7

Baker, R. S., Ocumpaugh, J. L., & Andres, J. M. A. L. (2020). BROMP quantitative field observations: A review. In R. Feldman (Ed.), *Learning science: Theory, research, and practice* (pp. 127–156). McGraw-Hill.

Bakhshinategh, B., Zaiane, O. R., ElAtia, S., & Ipperciel, D. (2018). Educational data mining applications and tasks: A survey of the last 10 years. *Education and Information Technologies*, *23*(1), 537–553. https://doi.org/10.1007/s10639-017-9616-z

Bannert, M. (2007). *Metakognition beim Lernen mit hypermedien* [Metacognition in learning with hypermedia]. Waxmann Verlag.

Ben-Eliyahu, A., & Bernacki, M. L. (2015). Addressing complexities in self-regulated learning: A focus on contextual factors, contingencies, and dynamic relations. *Metacognition and Learning*, *10*(1), 1–13. https://doi.org/10.1007/s11409-015-9134-6

Bernacki, M. L. (2018). Examining the cyclical, loosely sequenced, and contingent features of self-regulated learning: Trace data and their analysis. In D. H. Schunk & J. A. Greene (Eds.), *Handbook of self-regulation of learning and performance* (2nd ed.). Routledge. https://doi.org/10.4324/9781315697048-24

Bernacki, M. L., Yu, L., Kuhlmann, S. L., Plumley, R. D., Greene, J. A., Duke, R. F., Freed, R., Hollander-Blackmon, C., & Hogan, K. A. (2025). Using multimodal learning analytics to validate digital traces of self-regulated learning in a laboratory study and predict performance in undergraduate courses. *Journal of Educational Psychology, 117*(2), 176–205. https://doi.org/10.1037/edu0000890

Bergner, Y., Choi, I., & Castellano, K. E. (2019). Item response models for multiple attempts with incomplete data. *Journal of Educational Measurement*, *56*(2), 415–436. https://doi.org/10.1111/jedm.12214

Bishop, J., & Verleger, M. A. (2013, June). The flipped classroom: A survey of the research. In *2013 ASEE annual conference & exposition* (Paper 6219). American Society for Engineering Education. https://peer.asee.org/22585

Blikstein, P., & Worsley, M. (2016). Multimodal learning analytics and education data mining: Using computational technologies to measure complex learning tasks. *Journal of Learning Analytics*, *3*(2), 220–238. https://doi.org/10.18608/jla.2016.32.11

Bock, R. D., & Gibbons, R. D. (2021). *Item response theory*. John Wiley & Sons. https://doi.org/10.1002/9781119716723

Chalmers, R. P. (2012). Mirt: A multidimensional item response theory package for the R environment. *Journal of Statistical Software*, *48*(6), 1–29. https://doi.org/10.18637/jss.v048.i06

Chen, B., Knight, S., & Wise, A. F. (2018). Critical issues in designing and implementing temporal analytics. *Journal of Learning Analytics*, *5*(1), 1–9. https://doi.org/10.18608/jla.2018.53.1

Chi, M. T. H., & Wylie, R. (2014). The ICAP framework: Linking cognitive engagement to active learning outcomes. *Educational Psychologist*, *49*(4), 219–243. https://doi.org/10.1080/00461520.2014.965823

Dent, A. L., & Koenka, A. C. (2016). The relation between self-regulated learning and academic achievement across childhood and adolescence: A meta-analysis. *Educational Psychology Review*, *28*, 425–474. https://doi.org/10.1007/s10648-015-9320-8

Dong, G., & Pei, J. (2007). *Sequence data mining*. Springer.

Eddy, S. L., & Hogan, K. A. (2014). Getting under the hood: How and for whom does increasing course structure work? *CBE—Life Sciences Education*, *13*(3), 453–468. https://doi.org/10.1187/cbe.14-03-0050

Ericsson, K. A., & Simon, H. A. (1993). Protocol analysis: Verbal reports as data. MIT Press. https://doi.org/10.7551/mitpress/5657.001.0001

Fan, Y., Lim, L., van der Graaf, J., Kilgour, J., Raković, M., Moore, J., Molenaar, I., Bannert, M., & Gašević, D. (2022). Improving the measurement of self-regulated learning using multi-channel data. *Metacognition and Learning*, *17*(3), 1025–1055. https://doi.org/10.1007/s11409-022-09304-z

Fan, Y., van der Graaf, J., Lim, L., Raković, M., Singh, S., Kilgour, J., Moore, J., Molenaar, I., Bannert, M., & Gašević, D. (2022). Towards investigating the validity of measurement of self-regulated learning based on trace data. *Metacognition and Learning*, *17*(3), 949–987. https://doi.org/10.1007/s11409-022-09291-1

Fan, Y., Rakovic, M., van der Graaf, J., Lim, L., Singh, S., Moore, J., Molenaar, I., Bannert, M., & Gašević, D. (2023). Towards a fuller picture: Triangulation and integration of the measurement of self-regulated learning based on trace and think aloud data. *Journal of Computer Assisted Learning, 39*(4), 1303–1324. https://doi.org/10.1111/jcal.12801

Fan, H., Xu, J., Cai, Z., He, J., & Fan, X. (2017). Homework and students' achievement in math and science: A 30-year meta-analysis, 1986–2015. *Educational Research Review*, *20*, 35–54. https://doi.org/10.1016/j.edurev.2016.11.003

Ferguson, C. J. (2016). An effect size primer: A guide for clinicians and researchers. In A. E. Kazdin (Ed.), *Methodological issues and strategies in clinical research* (4th ed., pp. 301–310). American Psychological Association. https://doi.org/10.1037/14805-020

Franklin, J. (2005). The elements of statistical learning: Data mining, inference and prediction. *The Mathematical Intelligencer*, *27*(2), 83–85. https://doi.org/10.1007/BF02985802

Freijeiro-González, L., Febrero-Bande, M., & González-Manteiga, W. (2022). A critical review of LASSO and its derivatives for variable selection under dependence among covariates. *International Statistical Review*, *90*(1), 118–145. https://doi.org/10.1111/insr.12469

Friedman, J. H., Hastie, T., & Tibshirani, R. (2010). Regularization paths for generalized linear models via coordinate descent. *Journal of Statistical Software, 33*(1), 1–22. https://doi.org/10.18637/jss.v033.i01

Giannakos, M., Spikol, D., Di Mitri, D., Sharma, K., Ochoa, X., & Hammad, R. (2022). Introduction to multimodal learning analytics. In M. Giannakos, D. Spikol, D. Di Mitri, K. Sharma, X. Ochoa, & R. Hammad (Eds.), *The multimodal learning analytics handbook* (pp. 3–28). Springer Cham. https://doi.org/10.1007/978-3-031-08076-0_1

Greene, J. A., & Azevedo, R. (2010). The measurement of learners' self-regulated cognitive and metacognitive processes while using computer-based learning environments. *Educational Psychologist*, *45*(4), 203–209. https://doi.org/10.1080/00461520.2010.515935

Greene, J. A., Bernacki, M. L., & Hadwin, A. F. (2024). Self-regulation. In P. A. Schutz & K. R. Muis (Eds.), *Handbook of educational psychology* (4th ed., pp. 314–334). Routledge. https://doi.org/10.4324/9780429433726-17

Greene, J. A., Deekens, V. M., Copeland, D. Z., & Yu, S. (2018). Capturing and modeling self-regulated learning using think-aloud protocols. In D. H. Schunk & J. A. Greene (Eds.), Handbook of self-regulation of learning and performance (2nd ed., pp. 323–337). Routledge. https://doi.org/10.4324/9781315697048-21

Greene, J. A., Hutchison, L. A., Costa, L.-J., & Crompton, H. (2012). Investigating how college students' task definitions and plans relate to self-regulated learning processing and understanding of a complex science topic. *Contemporary Educational Psychology*, *37*(4), 307–320. https://doi.org/10.1016/j.cedpsych.2012.02.002

Gurung, A., Lin, J., Huang, Z., Borchers, C., Baker, R. S., Aleven, V., & Koedinger, K. R. (2025). *Starting seatwork earlier as a valid measure of student engagement*. arXiv. https://doi.org/10.48550/arXiv.2505.13341

Hallgren, K. A. (2012). Computing inter-rater reliability for observational data: An overview and tutorial. *Tutorials in Quantitative Methods for Psychology*, *8*(1), 23–34. https://doi.org/10.20982/tqmp.08.1.p023

He, Q., & von Davier, M. (2016). Analyzing process data from problem-solving items with n-grams: Insights from a computer-based large-scale assessment. In Y. Rosen, S. Ferrara, & M. Mosharraf (Eds.), *Handbook of research on technology tools for real-world skill development* (pp. 750–777). IGI Global. https://doi.org/10.4018/978-1-4666-9441-5.ch029

Hu, J., & Gao, X. (2017). Using think-aloud protocol in self-regulated reading research. *Educational Research Review*, *22*, 181–193. https://doi.org/10.1016/j.edurev.2017.09.004

Karabenick, S. A., Woolley, M. E., Friedel, J. M., Ammon, B. V., Blazevski, J., Bonney, C. R., De Groot, E., Gilbert, M. C., Musu, L., Kempler, T. M., & Kelly, K. L. (2007). Cognitive processing of self-report items in educational research: Do they think what we mean? *Educational Psychologist*, *42*(3), 139–151. https://doi.org/10.1080/00461520701416231

Kasakowskij, R., Haake, J. M., & Seidel, N. (2023). Self-assessment task processing behavior of students in higher education. In M. Feng, T. Käser, & P. Talukdar (Eds.), *Proceedings of the 16th international conference on educational data mining* (pp. 334–341). International Educational Data Mining Society. https://eric.ed.gov/?id=ED630881

Khosravi, H., Buckingham Shum, S., Chen, G., Conati, C., Tsai, Y.-S., Kay, J., Knight, S., Martinez-Maldonado, R., Sadiq, S., & Gašević, D. (2022). Explainable artificial intelligence in education. *Computers and Education: Artificial Intelligence*, *3*, Article 100074. https://doi.org/10.1016/j.caeai.2022.100074

Kim, H.-Y. (2017). Statistical notes for clinical researchers: Chi-squared test and Fisher's exact test. *Restorative Dentistry & Endodontics*, *42*(2), 152–155. https://doi.org/10.5395/rde.2017.42.2.152

Kirk, R. E. (1996). Practical significance: A concept whose time has come. *Educational and Psychological Measurement, 56*(5), 746–759. https://doi.org/10.1177/0013164496056005002

Kline, R. B. (2013). Beyond significance testing: Statistics reform in the behavioral sciences (2nd ed.). American Psychological Association. https://doi.org/10.1037/14136-000

Kraft, M. A. (2020). Interpreting effect sizes of education interventions. *Educational Researcher, 49*(4), 241–253. https://doi.org/10.3102/0013189X20912798

Kuhlmann, S. L., Plumley, R., Evans, Z., Bernacki, M. L., Greene, J. A., Hogan, K. A., Berro, M., Gates, K., & Panter, A. (2024). Students' active cognitive engagement with instructional videos predicts STEM learning. *Computers & Education*, *216*, Article 105050. https://doi.org/10.1016/j.compedu.2024.105050

Liao, C.-H., & Wu, J.-Y. (2023). Learning analytics on video-viewing engagement in a flipped statistics course: Relating external video-viewing patterns to internal motivational dynamics and performance. *Computers & Education*, *197*, Article 104754. https://doi.org/10.1016/j.compedu.2023.104754

Maher, J. M., Markey, J. C., & Ebert-May, D. (2013). The other half of the story: Effect size analysis in quantitative research. *CBE—Life Sciences Education*, *12*(3), 345–351. https://doi.org/10.1187/cbe.13-04-0082

Maldonado-Mahauad, J., Pérez-Sanagustín, M., Kizilcec, R. F., Morales, N., & Munoz-Gama, J. (2018). Mining theory-based patterns from Big data: Identifying self-regulated learning strategies in Massive Open Online Courses. *Computers in Human Behavior*, *80*, 179–196. https://doi.org/10.1016/j.chb.2017.11.011

Malekian, D., Bailey, J., & Kennedy, G. (2020). Prediction of students' assessment readiness in online learning environments: The sequence matters. In C. Rensing, H. Drachsler, V. Kovanović, N. Pinkwart, M. Scheffel, & K. Vertbert (Eds.), *LAK '20: Proceedings of the tenth international conference on learning analytics & knowledge* (pp. 382–391). ACM Press. https://doi.org/10.1145/3375462.3375468

McHugh, M. L. (2013). The Chi-square test of independence. *Biochemia Medica*, *21*(2), 143–149. https://doi.org/10.11613/BM.2013.018

Messick, S. (1989). Validity. In R. L. Linn (Ed.), *Educational measurement* (3rd ed., pp. 13–103). Macmillan Publishing.

Molenaar, I., & Järvelä, S. (2014). Sequential and temporal characteristics of self and socially regulated learning. *Metacognition and Learning*, *9*, 75–85. https://doi.org/10.1007/s11409-014-9114-2

Molenaar, I., & Wise, A. F. (2022). Temporal aspects of learning analytics: Grounding analyses in concepts of time. In C. Lang, G. Siemens, A. F. Wise, D. Gašević, & A. Merceron (Eds.), *The handbook of learning analytics* (pp. 66–76). Society for Learning Analytics Research. https://doi.org/10.18608/hla22.007

Moos, D. C., & Bonde, C. (2016). Flipping the classroom: Embedding self-regulated learning prompts in videos. *Technology, Knowledge and Learning, 21*, 225–242. https://doi.org/10.1007/s10758-015-9269-1

Nath, D., Gašević, D., Fan, Y., & Rajendran, R. (2024). CTAM4SRL: A consolidated temporal analytic method for analysis of self-regulated learning. In S. Joksimović & A. Zamecnik (Eds.), *LAK '24: Proceedings of the 14th learning analytics and knowledge conference* (pp. 645–655). ACM Press. https://doi.org/10.1145/3636555.3636926

Nicewander, W. A. (2018). Conditional reliability coefficients for test scores. *Psychological Methods*, *23*(2), 351–362. https://doi.org/10.1037/met0000132

Nystrand, M., Wu, L. L., Gamoran, A., Zeiser, S., & Long, D. A. (2003). Questions in time: Investigating the structure and dynamics of unfolding classroom discourse. *Discourse Processes*, *35*(2), 135–198. https://doi.org/10.1207/S15326950DP3502_3

Olson, S., & Riordan, D. G. (2012). Engage to excel: Producing one million additional college graduates with degrees in science, technology, engineering, and mathematics [Report to the President]. Executive Office of the President. https://eric.ed.gov/?id=ed541511

Osakwe, I., Chen, G., Fan, Y., Rakovic, M., Singh, S., Molenaar, I., & Gašević, D. (2024). Measurement of self-regulated learning: Strategies for mapping trace data to learning processes and downstream analysis implications. In S. Joksimović & A. Zamecnik (Eds.), *LAK '24: Proceedings of the 14th learning analytics and knowledge conference* (pp. 563–575). ACM Press. https://doi.org/10.1145/3636555.3636915

Pintrich, P. R., Smith, D. A. F., Garcia, T., & McKeachie, W. J. (1991). *A manual for the use of the motivated strategies for learning questionnaire (MSLQ)* [Technical Report No. 91-B-004]. National Center for Research to Improve Postsecondary Teaching and Learning. https://eric.ed.gov/?id=ED338122

Plumley, R. D., Bernacki, M. L., Greene, J. A., Kuhlmann, S., Raković, M., Urban, C. J., Hogan, K. A., Lee, C., Panter, A. T., & Gates, K. M. (2024). Co-designing enduring learning analytics prediction and support tools in undergraduate biology courses. *British Journal of Educational Technology*, *55*(5), 1860–1883. https://doi.org/10.1111/bjet.13472

Russell, D., & Winston, B. (2014). Tapping into the interpreting process: Using participant reports to inform the interpreting process in educational settings. *Translation & Interpreting*, *6*(1), 102–127. https://doi.org/10.12807/ti.106201.2014.a06

Salehian Kia, F., Hatala, M., Baker, R. S., & Teasley, S. D. (2021). Measuring students' self-regulatory phases in LMS with behavior and real-time self report. In M. Scheffel, N. Dowell, S. Joksimović, & G. Siemens (Eds.), *LAK21: 11th international learning analytics and knowledge conference* (pp. 259–268). ACM Press. https://doi.org/10.1145/3448139.3448164

Sebesta, A. J., & Speth, E. B. (2017). How should I study for the exam? Self-regulated learning strategies and achievement in introductory biology. *CBE—Life Sciences Education*, *16*(2), Article 30. https://doi.org/10.1187/cbe.16-09-0269

Seo, K., Dodson, S., Harandi, N. M., Roberson, N., Fels, S., & Roll, I. (2021). Active learning with online video: The impact of learning context on engagement. *Computers & Education*, *165*, Article 104132. https://doi.org/10.1016/j.compedu.2021.104132

Shekhar, P., Borrego, M., DeMonbrun, M., Finelli, C., Crockett, C., & Nguyen, K. (2020). Negative student response to active learning in STEM classrooms. *Journal of College Science Teaching*, *49*(6), 45–54. https://doi.org/10.1080/0047231X.2020.12290664

Siegler, R. S., & Crowley, K. (1991). The microgenetic method: A direct means for studying cognitive development. *American Psychologist*, *46*(6), 606–620. https://doi.org/10.1037/0003-066X.46.6.606

Siemens, G. (2013). Learning analytics: The emergence of a discipline. *American Behavioral Scientist*, *57*(10), 1380–1400. https://doi.org/10.1177/0002764213498851

Theobald, E. J., Hill, M. J., Tran, E., Agrawal, S., Arroyo, E. N., Behling, S., Chambwe, N., Cintrón, D. L., Cooper, J. D., Dunster, G., Grummer, J. A., Hennessey, K., Hsiao, J., Iranon, N., Jones, L., II., Jordt, H., Keller, M., Lacey, M. E., Littlefield, C. E., … Freeman, S. (2020). Active learning narrows achievement gaps for underrepresented students in undergraduate science, technology, engineering, and math. *Proceedings of the National Academy of Sciences*, *117*(12), 6476–6483. https://doi.org/10.1073/pnas.1916903117

Ulitzsch, E., He, Q., & Pohl, S. (2022). Using sequence mining techniques for understanding incorrect behavioral patterns on interactive tasks. *Journal of Educational and Behavioral Statistics*, *47*(1), 3–35. https://doi.org/10.3102/10769986211010467

Ulitzsch, E., Ulitzsch, V., He, Q., & Lüdtke, O. (2023). A machine learning-based procedure for leveraging clickstream data to investigate early predictability of failure on interactive tasks. *Behavior Research Methods*, *55*(3), 1392–1412. https://doi.org/10.3758/s13428-022-01844-1

Weinberg, A., & Thomas, M. (2018). Student learning and sense-making from video lectures. *International Journal of Mathematical Education in Science and Technology, 49*(6), 922–943. https://doi.org/10.1080/0020739X.2018.1426794

Wickham, H. (2023). *stringr: Simple, consistent wrappers for common string operations* (Version 1.5.0) [Computer software]. Posit Software. https://CRAN.R-project.org/package=stringr

Wiedbusch, M., Dever, D., Li, S., Amon, M. J., Lajoie, S., & Azevedo, R. (2023). Measuring multidimensional facets of SRL engagement with multimodal data. In V. Kovanovic, R. Azevedo, D. C. Gibson, & D. Ifenthaler (Eds.), *Unobtrusive observations of learning in digital environments: Examining behavior, cognition, emotion, metacognition and social processes using learning analytics* (pp. 141–173). Springer Cham. https://doi.org/10.1007/978-3-031-30992-2_10

Winne, P. H. (2017). Learning analytics for self-regulated learning. In C. Lang, G. Siemens, A. Wise, D. Gašević (Eds.), *Handbook of learning analytics* (pp. 241–249). Society for Learning Analytics Research. https://doi.org/10.18608/hla17.021

Winne, P. H. (2020). Construct and consequential validity for learning analytics based on trace data. *Computers in Human Behavior, 112*, Article 106457. https://doi.org/10.1016/j.chb.2020.106457

Winne, P. H. (2022). Learning analytics for self-regulated learning. In C. Lang, G. Siemens, A. F. Wise, D. Gašević, & A. Merceron (Eds.), *The handbook of learning analytics* (pp. 78–85). Society for Learning Analytics Research. https://doi.org/10.18608/hla22.008

Winne, P. H., & Hadwin, A. F. (1998). Studying as self-regulated learning. In D. J. Hacker, J. Dunlosky, & A. C. Graesser (Eds.), *Metacognition in educational theory and practice* (pp. 277–304). Lawrence Erlbaum Associates.

Winne, P. H., & Hadwin, A. F. (2008). The weave of motivation and self-regulated learning. In D. H. Schunk & B. J. Zimmerman (Eds.), *Motivation and self-regulated learning: Theory, research, and applications* (pp. 297–314). Lawrence Erlbaum Associates.

Winne, P. H., Jamieson-Noel, D., & Muis, K. (2002). Methodological issues and advances in researching tactics, strategies, and self-regulated learning. In P. R. Pintrich & M. L. Maehr (Eds.), *New directions in measures and methods* (pp. 121–155). Advances in motivation and achievement volume 12. Emerald Publishing.

Wise, A. F., & Shaffer, D. W. (2015). Why theory matters more than ever in the age of big data. *Journal of Learning Analytics*, *2*(2), 5–13. https://doi.org/10.18608/jla.2015.22.2

Zhang, J., Borchers, C., Aleven, V., & Baker, R. S. (2024). Using large language models to detect self-regulated learning in think-aloud protocols. In B. Paaßen & C. D. Epp (Eds.), *Proceedings of the 17th international conference on educational data mining* (Paper 13). International Educational Data Mining Society. https://doi.org/10.5281/zenodo.12729790

Zhang, Y., & Paquette, L. (2023). Sequential pattern mining in educational data: The application context, potential, strengths, and limitations. In A. Peña-Ayala (Ed.), *Educational data science: Essentials, approaches, and tendencies* (pp. 219–254). Springer Singapore. https://doi.org/10.1007/978-981-99-0026-8_6

Zimmerman, B. J. (1986). Becoming a self-regulated learner: Which are the key subprocesses? *Contemporary Educational Psychology*, *11*(4), 307–313. https://doi.org/10.1016/0361-476X(86)90027-5

Zou, H., & Hastie, T. (2005). Regularization and variable selection via the elastic net. *Journal of the Royal Statistical Society Series B: Statistical Methodology, 67*(2), 301–320. https://doi.org/10.1111/j.1467-9868.2005.00503.x

Appendix

Classes*, descriptions, and examples of the macro- and micro-level SRL processes used to code students' self-regulatory behaviour. Micro-level SRL processes are grouped according to their superordinate macro-level SRL variable.*

Micro-level processes	Description[a]	Student example
Macro-Level Variable: Task Codes		
Planning (PLAN)	Stating two or more sub-goals simultaneously or stating a sub-goal and combining it with a time requirement.	"I'm going to watch I'm writing down on my scratch paper watch video as number one. And then number two is answer the Learning Catalytics question."
Sub-goal (SG)	Learner articulates a specific sub-goal that is relevant to the experiment provided overall goal.	"I'm going to rewind to figure out why I'm doing this."
Recycle goal in working memory (RECYCLE)	Restating the goal (e.g., question or parts of a question) in working memory.	"I'm just going to review this again real quick. So, lesson, watch Ellipses video, and then complete…"
Task definition (TDFN)	Identifying the prompt or task required of the learner by the assignment or lesson.	"I'm going to check the assignment list again. I don't remember. Okay. So, it's... Watch the videos first and then complete the homework assignment."
Macro-level variable: Domain-General Strategies		
Comparing and contrasting (CC)	Examining representations of *two or more different concepts/ideas* to determine how they are similar or different, with a focus on gaining understanding of them.	"I see that they got the same answer as me I think we went about it a different way so they grouped their X squared minus 4X and then they did 3Y squared plus 24Y equals 29 and then they did oh okay so they did X squared minus 4X plus 4 so and then that's how they did that."
Coordinating informational sources (COIS)	Comparing two representations to develop understanding of a single concept.	NA
Draw (DRAW)	Making a drawing or diagram to assist in learning.	"I'm just going to quickly make a graph of that and plot it. And then it says endpoint (6, 2), so I'm going to plot that five, six, six, two."
Emotion regulation (ER)	Actively attempting to control emotional response to some aspect of the learning task.	NA
FNC	Putting together two pieces of information and drawing a new conclusion that extends beyond what is presented in the hypermedia environment.	"I realized what I did here. So, the minor axis was the one that's basically usually underneath the foci, but the major axis is where the foci lie on so that was my bad."
HSB	Learner seeks assistance regarding either the adequacy of their understanding or their learning behavior.	"I feel like I would just stop now and then go to office hours."
Inferring source content (ISC)	Learner makes a guess as to the content available in a source while observing hyperlink, title, or description. Must occur before significantly engaging in content.	"but I'm assuming that will probably be discussed later in the video."
Memorization (MEM)	Learner tries to memorize text, diagram, etc.	NA
Prior knowledge activation (PKA: Course)	Learner retrieves prior knowledge (i.e., declarative, procedural, and/or conceptual) from memory.	"Usually these things are 15 questions long and it takes forever."
Prior knowledge activation (PKA: Content)	Learner retrieves prior knowledge (i.e., declarative, procedural, and/or conceptual) from memory.	"I remember her saying that you have to add the things over, so…"
Read notes (RN)	Learner reads over their own notes, drawings, etc.	"I'll look through my notes and try to find where she talks about the vertex."

Rereading (RR)	Rereading or revisiting a section of the hypermedia environment.	"I'm going back a little bit in the video to reorient myself."
Search (SEARCH)	Searching the hypermedia environment.	"So, I'm going to look up ellipse a b c relationship."
Self-questioning (SQ)	Learner asks a specific, substantive question relevant to the task/content.	"I wonder why H and K are being used instead of X and Y."
Self-testing (ST)	Learner formulates a question (SQ) or reads a question in the hypermedia environment that is outside of formative assessment (i.e., questions that are not a part of quizzes, GRQs, or learning catalytics), then verbalizes an answer to the question without reading answer explicitly from environment.	"Why didn't she factor out the 16? In order to be able to complete the square easier, I guess."
Summarization (SUM)	Verbally restating what was just read, inspected, or heard in the hypermedia environment.	"I think student one was right in theory but just failed to see that the major axis is the X, making it a horizontal ellipse. But that's a mistake that can easily be corrected."
Taking notes (TN)	Learner writes down information.	"I'm just going to write that down for future reference. The foci is called... Wait the line containing the foci is called the major axis."
Macro-Level Variable: Monitoring		
Content evaluation (plus or minus; CE+/−)[b]	Monitoring content relevant to goals. Learner states content is or is not useful toward reaching the goal.	+: "That's a good example." -: "That was a little bit confusing in terms of the explanation."
Emotion monitoring (EM)	Learner realizes that they are having an emotional response due to some aspect of the learning task.	"I'm a little overwhelmed."
Expectation of adequacy of content (plus or minus; EAC +/−)[b]	Expecting that a certain type of representation will prove either adequate or inadequate given the current goal.	+: "NA" -: "I don't really care about the extra examples about elementary school or if she's giving an example." [EAC-]
Feeling of knowledge (plus or minus; FOK +/−)[b]	Learner is aware of having read something in the past and having some understanding of it, but not being able to recall it on demand or learner states this is information not seen before.	+: "I know how to do this now, so maybe this will work." -: "I don't remember how to determine which one is going to be the minor or major."
Feeling of recognition (plus or minus; FOR +/−)[b]	Participant sees a representation and remembers encountering it in the past.	+: "I remember from before homework." -: "I don't think she covered this in the video."
Judgment of correct (plus or minus; JOC +/−)[b]	When a learner makes a metacognitive judgment that *their* answer to a question presented in the learning environment (e.g., checkpoint question during ST) is correct (JOC+) or incorrect (JOC-).	+: [in reference to an answer] "I think that's right." -: [in reference to an answer] "I'm not totally sure…"
Judgment of learning (plus or minus; JOL +/−)[b]	A judgment that what was just learned will be remembered at a future time, such as the post-test.	+: "." -: "because I feel like I didn't really absorb what she said right there."
Judgment of understanding (plus or minus; JOU +/−)[b]	A judgment that what was just read/viewed/heard/etc. was sufficiently understood.	+: "I get it." -: "That doesn't make sense."
Monitoring information coherence (plus or minus; MIC +/−)[b]	Learner notes that a recently viewed piece of information agrees with information they already encountered whether in the current learning environment or from prior knowledge.	+: "So that's basically the same thing the other one said." -: "which is not what they did."
Monitoring math accuracy (MMA)	Assessing one's use of Mathematical Problem-Solving (MPS), one's own quality of work within MPS, or any	"Okay. What? Oh, wait. No."

ISSN 1929-7750 (online).

	products of MPS. MMA is a special case of JOU for math calculations.	
Monitor progress (MP)	Assessing whether participant's previously set goal and/or participant's own standard for understanding has been met.	"I did my before class, and that's homework, so I think I'm done."
Monitor use of strategies (plus or minus; MUS+/−)	Participant comments on how useful a strategy is/was.	+: "NAit" -: "And that's why I hate using the computer calculator."
Monitoring understanding (MU)	Evaluating the appropriateness of one's understanding of the goal or the task and the possible pathways to achieving those goals. This answers the question: am I doing what I should be doing? Note that the task or goal can be the learning prompt or the specific questions, if there are questions in the task.	"No, this is a video I just watched. So I'm going to give it back now. Oh, okay. That's definitely what I need."
Self-knowledge activation (SKA)	The participant verbalizes that they are going to invoke a strategy because it is helpful to him/her personally. Or participant verbalizes that he/she is NOT going to invoke a strategy because it is NOT helpful to him/her. Or, participant says something about his/her own knowledge, beliefs, disposition, etc.	"I'm not the greatest at factoring."
Task difficulty (TD)	Learner indicates one of the following: 1) the task is either easy or difficult, 2) the questions are either simple or difficult, 3) using the hypermedia environment is easier or more difficult than using a book.	"And then we're going to do the focus, which is a little difficult."
Time monitoring (TM)	Participant refers to the numbers of minutes remaining.	"I'm going to check to see how much time I have: seven minutes."
Macro-Level Variable: Domain-Specific Strategies		
Mathematical problem-solving (MPS)	Learner works through math problem (e.g., quiz question, learning catalytics) to answer a question.	"I'll have x squared +4x. Then I'll have 4y squared -24y in parentheses, will equal -4."
Macro-Level Variable: Assessment Strategies		
Changing answers	Intentionally returning to a question to change your answer.	NA
Match	A low-level strategy where the learner searches learning environment for content specific to a learning task prompt (e.g., homework question, GRQ).	"I'll just skim through the other video until I find something that has to do with the vertex."
Guessing	Selecting a multiple-choice answer option as a guess rather than because they know it's the answer.	"I guess I'm just going to have to guess, because I really don't know."
Ruling out answers	Reviewing answer choices on a multiple-choice question and systematically ruling out answer choices in order to narrow down to the answer they will select.	"So that knocks it down to C or D."
Skipping an item	A participant decides to skip an item on an assignment or task (e.g., quiz item), whether they will come back to it later or not.	"So I might skip this and come back."
Testwiseness	Using the multiple-choice format, feedback, or other structural qualities of the test to figure out what an answer is.	"I'm going to say ellipse because that is what we are working with."