Subtopic-specific Heterogeneity in Computer-based Learning Behaviors

HaeJin Lee[1], Nigel Bosch[1,2]

[1]School of Information Sciences, University of Illinois Urbana–Champaign, USA
[2]Department of Educational Psychology, University of Illinois Urbana–Champaign, USA

Author Note

Correspondence concerning this article should be addressed to HaeJin Lee, School of Information Sciences, 501 E. Daniel St., Champaign, IL 61820, USA. Email: haejin2@illinois.edu

Declarations

Ethics approval and consent to participate

This research was reviewed by Institutional Review Board (IRB) and all participants completed an IRB-approved consent form (IRB protocol #21019).

Consent for publication

Consent for publication has been obtained from all participants.

Availability of supporting data/Availability of data and materials

The datasets used and/or analyzed during the current study are available from the corresponding author on reasonable request. The software used for data collection and analysis is available online (https://osf.io/j9h74/?view_only=a93f7b3649414b288933cc73fb188795).

Competing interests

The authors have no competing interests to disclose.

Funding

This research was supported by NSF grant no. 2202481. Any opinions, findings, and conclusions or recommendations expressed in this material are those of the author(s) and do not necessarily reflect the views of the National Science Foundation.

Authors' contributions

HL and NB contributed to the conceptualization, software development, and methodology of this research. HL wrote the original draft, conducted formal analysis, data curation, and visualization. NB reviewed and edited the manuscript.

Acknowledgements

We sincerely appreciate the reviewers for providing valuable comments and suggestions, which helped improve this manuscript.

Abstract

Background

Self-regulated learning (SRL) strategies can be domain specific. However, it remains unclear whether this specificity extends to different subtopics within a single subject domain. In this study, we collected data from 210 college students engaged in a computer-based learning environment to examine the heterogeneous manifestations of learning behaviors across four distinct subtopics in introductory statistics. Further, we explore how the time spent engaging in metacognitive strategies correlated with learning gain in those subtopics.

Results

By employing two different analytical approaches that combine data-driven learning analytics (i.e., sequential pattern mining in this case), and theory-informed methods (i.e., coherence analysis), we discovered significant variability in the frequency of learning patterns that are potentially associated with SRL-relevant strategies across four subtopics. In a subtopic related to calculations, engagement in coherent quizzes (i.e., a type of metacognitive strategy) was found to be significantly less related to learning gains compared to other subtopics. Additionally, we found that students with different levels of prior knowledge and learning gains demonstrated varying degrees of engagement in learning patterns in an SRL context.

Conclusion

The findings imply that the use—and the effectiveness—of learning patterns that are potentially associated with SRL-relevant strategies varies not only across contexts and domains, but even across different subtopics within a single subject. This underscores the importance of personalized, context-aware SRL training interventions in computer-based learning environments, which could significantly enhance learning outcomes by addressing the heterogeneous relationships between SRL activities and outcomes. Further, we suggest theoretical implications of subtopic-specific heterogeneity within the context of various SRL models. Understanding SRL heterogeneity enhances these theories, offering more nuanced insights into learners' metacognitive strategies across different subtopics.

Keywords: Self-regulated learning, Learning analytics, Heterogeneity.

Introduction

Computer-based learning environments offer a flexible and adaptive learning experience, granting students significant autonomy. However, these environments also present distinct challenges, particularly for students who have not yet acquired all the necessary self-regulated learning (SRL) skills (Bol & Garner, 2011; Irfan et al., 2020; Pedrotti & Nistor, 2019). SRL is a learner's active management and adaptation of their learning strategies to meet their learning goals and overcome challenges encountered throughout learning. Students with SRL skills possess the ability to orchestrate their learning plans strategically, as well as the capacity to reflect upon and assess their learning progress continually, which ultimately benefits learning (Azevedo, 2005; Johnson et al., 2011; van Alten et al., 2020; van der Graaf et al., 2022). Therefore, the inherent freedom and complexity of computer-based learning environments, although beneficial in numerous ways, often pose challenges in navigation and success, particularly for learners who are still developing their SRL skills (Taub et al., 2021; Zheng et al., 2022).

Although SRL is crucial for effective learning, it is not an inherent skill among students and varies substantially between students (Bernacki et al., 2015; Muwonge et al., 2020). This variability in SRL skills often reflects disparities in educational resources and learning opportunities, rather than an inherent flaw or lack of potential in the students themselves (Zimmerman, 2002). Fortunately, however, SRL skills are not static but can be developed and enhanced over time (Zimmerman & Kitsantas, 2005). Thus, there are promising opportunities for interventions to teach SRL skills; although some students may not have had sufficient opportunities to hone these skills, SRL can be progressively learned and enhanced with appropriate guidance and practice (Bernacki et al., 2020; Schunk & Zimmerman, 1997; Zimmerman, 2002).

While SRL-supporting tools are crucial in fostering students' SRL skills (Bellhäuser et al., 2023; T. Li et al., 2023), current approaches tend to model and support a uniform application of SRL strategies across various learning domains and subdomains. Broadbent et al. (2020) highlight a challenge in the development of SRL interventions, questioning whether it is more effective to design these SRL interventions with a focus on specific domains or to apply a more general approach across domains. While Broadbent and colleagues acknowledge that non-content-specific SRL strategies can be beneficial, they discuss that content-specific approaches to SRL interventions might be more effective. In line with this perspective, numerous studies confirmed that SRL strategies are not one-size-fits-all and are indeed subject-specific and underscored the need for domain-specific SRL approaches, since they are not universally applicable (Alexander et al., 2011; Greene, Bolick, Caprino, et al., 2015; Greene, Bolick, Jackson, et al., 2015; Lee et al., 2023; Poitras & Lajoie, 2013). However, such evidence prompts further questions about whether such specificity in SRL strategies should extend to the varied subtopics within a single subject domain, such as mathematics, computer science, and humanities.

Within subject domains, there are more narrowly focused areas, which we refer to as subtopics, that potentially demand unique problem-solving approaches. For instance, in this study, we refer to statistics as a subject domain, which involves the study of collecting, interpreting, and analyzing data. Even within the subject of statistics, there exist numerous subtopics, such as calculation and graph interpretation, each requiring distinct problem-solving methods. For example, calculation involves tasks like probability computations and standard deviation calculations, which rely on direct application of mathematical formulas. In contrast, graph interpretation entails understanding graphical data presented in formats such as scatter plots or histograms, demanding different skills. Similarly, in computer science, the subtopic of programming requires an understanding of the syntax and semantics of various languages, along with coding skills, while the data structures subtopic demands a deep understanding of algorithms, including sorting and search algorithms.

For students who struggle with SRL skills, recognizing and adapting the appropriate cognitive and metacognitive strategies to the specific demands of each subtopic poses a significant challenge. A generalized approach may not sufficiently account for the intricate variances in how SRL strategies are employed (and should be employed), even across different subtopics within a single domain. Therefore, there is a need to develop AI-based systems that can support students' personalized learning by fostering the development of SRL skills tailored to specific subtopics in computer-based learning environments. Numerous studies explored the heterogeneous application of SRL strategies across diverse student populations, taking into account variables such as gender, race, and academic performance (Carroll & Garavalia, 2002; Foong et al., 2021; Norman & Furnes, 2016; Virtanen & Nevgi, 2010; Yukselturk & Bulut, 2009; Zimmerman & Pons, 1990). However, there remains a gap in understanding how SRL strategies can and should vary across different subtopics within a single domain in SRL research.

In this paper, we explore this issue by investigating differential engagement in SRL-relevant learning patterns across four different subtopics within the subject domain of statistics. To the best of our knowledge, this study is the first to explore the heterogeneity of computer-based learning behaviors in an SRL context across various subtopics within a single domain. While SRL skills benefit many academic outcomes, in general, and in domain-specific research

(Kramarski & Gutman, 2006; Mason et al., 2010; Schraw et al., 2006; Tseng et al., 2006), our exploration challenges the conventional belief that increased SRL engagement invariably leads to higher learning gains, irrespective of the subtopics. Our focus is on discerning whether the correlation between SRL engagement and learning gains is consistent across various subtopics or shows notable variations. Such investigations are crucial as they question the generalizability of the efficacy of SRL strategies and offer insights into how SRL skills could be taught in a more targeted, effective fashion. By exploring the complexities of SRL heterogeneity, our study aims to make contributions in two ways. First, we expect that the insights gained from our study will contribute to enhancing the development of more personalized and context-dependent AI-based systems, thereby enhancing the overall effectiveness of SRL in computer-based learning environments. Second, we anticipate that our findings will enrich existing SRL theories by revealing the potential to account for variations in SRL-relevant strategies based on specific subtopics.

Our research is structured into two closely related analyses. Analysis 1 employs a data-driven approach to explore the heterogeneity of learning patterns in an SRL context across subtopics, addressing two specific research questions (RQ1 and RQ2) using sequential pattern mining. Analysis 2, which addresses RQ3, takes a theory-driven approach to examine the heterogeneous relationship between time spent using metacognitive strategies (a type of SRL skill) and learning gains (measured as the difference between posttest and pretest grades) across subtopics. While Analysis 1 focuses on uncovering the varied nature of learning patterns in an SRL context—questioning whether the frequency of employing specific learning patterns differs based on the subtopic—Analysis 2 advances this inquiry by examining the extent to which metacognitive strategies produced comparable learning gains across different subtopics.

Our research questions were as follows:

RQ1. Are there variations in the frequency of learning patterns in an SRL context across different subtopics?

RQ2. How does the association between learning gain (measured as the difference between posttest and pretest grades) and the frequency of learning patterns in an SRL context vary across different subtopics? Furthermore, how does the association between prior knowledge (measured as the pretest grade) and the frequency of learning patterns that are potentially associated with SRL-relevant strategies vary across different subtopics?

RQ3. Does the relationship between time spent on metacognitive strategy use and learning gain vary across different subtopics?

Theoretical Models of SRL

SRL is a comprehensive framework that includes cognitive, metacognitive, affective, and behavioral facets of learning (Panadero, 2017; Schunk & Greene, 2017). Numerous theoretical models have been developed to understand SRL (Efklides, 2011; Pintrich, 2000; Zimmerman, 1989; Zimmerman & Moylan, 2009), with several specifically designed to subdivide and categorize the processes inherent to SRL. For instance, Zimmerman's SRL model (Zimmerman & Moylan, 2009) comprises three phases: forethought, performance, and self-reflection. In the forethought phase, students engage in preparatory steps, including analyzing the task, setting goals, and planning their strategies, to establish a foundation for their learning process. In the performance phase, students execute the learning strategies by managing time, monitoring their progress, and using metacognitive strategies to keep themselves motivated. Lastly, in the self-reflection phase, students assess and reflect on their goals, strategies, and plans to set their future learning.

In Winne and Hadwin's SRL model (Winne & Hadwin, 1998), which has a strong focus on metacognition, students actively manage their learning by monitoring and employing (meta)cognitive strategies. Specifically, this model highlights the goal-driven nature of SRL and the impact of self-regulatory actions on motivation. Winne and Hadwin's model also provides a detailed examination of the interaction between various SRL components. The model acknowledges that SRL occurs across phases but differs from many other models by also modeling the information processes that occur within each phase (Azevedo et al., 2010; Winne & Hadwin, 1998). Based on Winne and Hadwin's model, students employ five distinct facets—conditions, operations, products, evaluations, and standards—within tasks that unfold over four phases. These phases include task definitions, goal setting, the enactment of study tactics, and metacognitive adaptations to studying. Although there exist differences within a multitude of SRL models (Efklides, 2011; Pintrich, 2000; Winne & Hadwin, 1998; Zimmerman & Moylan, 2009), especially in terms of the focus of the model and perspective the researchers are using to understand SRL process (e.g., Zimmerman uses a socio-cognitive perspective of SRL and Winne and Hadwin use the view of information processing theory), researchers agree that SRL consists of different phases and subprocesses that students revisit repeatedly throughout learning. Further, one common facet of SRL models is the use of metacognitive strategies during learning. Using metacognitive strategies, such as task analysis, goal setting, selecting and applying strategies, and monitoring and reflection on learning, are key components across many SRL models (Panadero, 2017; Puustinen & Pulkkinen, 2001; Schunk & Greene, 2017).

Supporting SRL in Computer-based Online Learning Environments

Investigating heterogeneity in an SRL context is particularly important not only because it provides opportunities to observe the array of strategies students use to steer their own learning, but also because it pinpoints areas where students may benefit from additional support or instruction regarding SRL skills in computer-based learning environments. Research demonstrated the critical role of SRL in online learning environments, showing a positive relationship between employing these strategies and academic achievement (Jin et al., 2023; Johnson et al., 2011; Richardson et al., 2012; Xu et al., 2023). However, computer-based learning environments often demand higher levels of SRL skills compared to traditional in-person courses, as students are required to independently monitor their learning processes and make continuous adjustments as necessary. For instance, students must decide when and how to engage with the course content, often with minimal guidance beyond the course's structural design (Lajoie & Azevedo, 2006).

This autonomy underscores the necessity for students to exhibit a significant capacity for SRL skills to achieve the required learning objectives (Artino & Stephens, 2009; Barnard et al., 2008; Broadbent & Poon, 2015; Kizilcec & Schneider, 2015). Therefore, providing individualized support could be especially beneficial to students who lack SRL skills since those students often confront challenges in navigating and succeeding within these learning environments (Aleven & Koedinger, 2002; Graesser & McNamara, 2010; Greene et al., 2010). In response to this, numerous studies attempted to foster and support students' SRL skills in online learning environments through a variety of approaches. These methods include open learner models (Bull et al., 2014; Ferreira da Rocha et al., 2023; Guerra et al., 2016, 2018; Kay et al., 2022; Law et al., 2017; Sun et al., 2023; Tacoma et al., 2018; Winne, 2021), dashboards (Alphen

& Bakker, 2016; Hsiao et al., 2016; Mejia et al., 2017; Muldner et al., 2015), interventions (Cicchinelli et al., 2018; Jansen et al., 2020; Müller & Seufert, 2018; Zarei Hajiabadi et al., 2023), metacognitive prompts (Engelmann et al., 2021; Pieger & Bannert, 2018; Sonnenberg & Bannert, 2019), and others. For systematic literature reviews of SRL-supporting tools, see Alvarez et al. (2022), Araka et al. (2020), Edisherashvili et al. (2022), Heikkinen et al. (2023), Hooshyar et al. (2020), and Matcha et al. (2020). Although tools supporting SRL are crucial for enhancing students' use of SRL skills, existing methods usually adopt a one-size-fits-all approach to SRL support, even across subtopics within various domains.

Moreover, among the tools designed to support students' SRL skills or behaviors, only a few studies utilized recommendations on which specific SRL skills should be used to actively guide students in developing their SRL capabilities (Du & Hew, 2022). For instance, Bodily et al. (2018) developed a content recommender which aimed to support students identifying knowledge gap by providing them summary of their mastery level of each concept. Additionally, they designed a skill recommender that provides students with an overview of their metacognitive strategy use, along with corresponding recommendations to support students' application of these strategies in an introductory blended chemistry course, at the university level. While Bodily et al. (2018) found that the majority of students who received the recommendations had positive feedback, these SRL strategy recommendations could be further personalized by suggesting effective strategies tailored to each subtopic in the course. Despite advancements in SRL-supporting tools, there is still significant potential for these tools or AI-based systems to offer students more personalized, content-dependent SRL support depending on the subtopics.

Measuring and Analyzing SRL through a Temporal Perspective

The effectiveness of SRL support is contingent upon the accurate measurement of students' SRL skills and SRL-related behaviors in computer-based learning environments (Q. Li et al., 2020; Winters et al., 2008). However, measuring SRL skills and behaviors is a multifaceted challenge (Greene & Azevedo, 2010; Hadwin et al., 2007; Winne, 2010; Winne & Perry, 2000). Researchers suggest SRL measures should be viewed as both aptitudes and events (Bannert et al., 2014; Winne, 2010). The aptitude-based approach focuses on students' characteristics, such as their cognitive, motivational, and emotional dispositions, and how these affect their ability to regulate their learning, treating SRL as a set of relatively static traits. This approach often uses questionnaires and structured interviews (Bannert et al., 2014; Winne, 2010). Some of the most used questionnaires and structured interviews include the Motivated Strategies for Learning Questionnaire (MSLQ) (Pintrich et al., 1991), the Self-Efficacy for Learning Form (SELF) (Zimmerman & Kitsantas, 2007), and the Online Self-Regulated Learning Questionnaire (OSLQ) (Barnard et al., 2008, 2010). However, despite its widespread use, the aptitude-based approach has been critiqued for portraying SRL as a fixed trait (Azevedo, 2015; Veenman & van Cleef, 2019).

Moreover, the reliance on students' perceptions and memories in questionnaires may not accurately reflect their *in situ* behaviors and strategies in a learning situation. As Greene and Azevedo (2010) argue, the aptitude-based approach can be incomplete since it does not account for the dynamic nature of SRL, wherein learners continuously adapt their learning processes within and between tasks in response to the unique demands of each. Similarly, trace data—digital footprints that learners leave behind as they interact with online learning environments, such as clicking on a link, submitting an answer, or spending time on a page (Brusilovsky, 2001; Du et al., 2023)—also has inherent limitations in capturing learner's self-perceptions. For instance, Choi et al. (2023) found substantial differences between students' self-reported goals and their goal-relevant behaviors reflected in trace data. However, this substantial misalignment indicates that trace data can serve as a counterpoint to self-perception measures. While there exist limitations in capturing SRL comprehensively using trace data alone, numerous studies have highlighted the discrepancies between self-reported and trace data, demonstrating the value of trace data in providing objective insights into student behaviors (Choi et al., 2023; Hadwin et al., 2007; F. Han, 2023; Syal & Nietfeld, 2020; Winne & Jamieson-Noel, 2002). For instance, studies discovered that student's trace data were better predictors of student game performance and academic performance than self-reported data (Syal & Nietfeld, 2020; Ye & Pennisi, 2022). Likewise, studies increasingly rely on trace data because it captures actual student actions in real-time, which could reduce the biases and inaccuracies often associated with self-reports (Palanci et al., 2024).

Given the limitations of the aptitude-based approach to SRL measurement, researchers have shifted towards examining SRL as a dynamic, temporal process (Fan et al., 2021; Saint et al., 2021). Process models, for example, focus on students' self-regulatory actions within specific contexts or tasks, viewing SRL as an unfolding series of actions and decisions in response to specific task demands (Cloude et al., 2022; Hardy et al., 2019; Klug et al., 2011; Winne & Perry, 2000). The process-based perspective has opened new ways to explore SRL but also introduced new challenges (Molenaar et al., 2023). First, shifting to a temporal perspective requires innovative methods for conceptualizing SRL's multi-dimensionality and dynamic nature (Azevedo, 2014; Järvelä et al., 2019; Jovanović et al., 2017). Reimann (2009) suggest that the temporal conceptualization of SRL should extend beyond mere time-on-task, frequency, and duration to also include the sequential order of learning events. Despite challenges with interpreting temporality and choosing measurement units, numerous studies investigated SRL as a series of events over time to better understand its dynamic nature. For instance, Maldonado-Mahauad et al. (2018) conceptualized SRL measurements by using questionnaire and process mining to extract students' learning interaction in massive open online courses. They identified six different interaction sequence patterns and related each pattern with corresponding SRL strategies grounded in literature. Although the authors further discuss the challenges that emerge while extracting theory-based patterns from observed behaviors, their study advances SRL research by providing a deeper understanding of how students engage with course content and assessments through the identification of SRL strategies in massive open online courses.

The second challenge stems from handling complex trace data, which demands advanced analytical techniques to extract meaningful insights into students' use of SRL skills (Gašević et al., 2015; Kizilcec et al., 2017; Siemens & Baker, 2012). In response to this challenge, numerous temporally focused learning analytics methods to measure and analyze SRL emerged, each with unique strengths and potential limitations. One method employed is lag-sequential analysis, as Kuvalja et al. (2014) used. This technique is used to examine the timing and order of events, with an emphasis on investigating the timing of actions, which can be beneficial for understanding the connections between events. Methods such as process mining and epistemic network analysis have been applied for a more holistic view of the SRL process. Process mining is a technique used to analyze and visualize sequences of processes based on event logs (Bogarín et al., 2018; Saint et al., 2021; Sobocinski et al., 2017). Despite its limitation of not allowing a global statistical test for group differences and varying individual weights, process mining can offer a detailed view of the sequence and flow of SRL events. Meanwhile, epistemic network analysis, which is grounded in epistemic frames theory (Shaffer, 2004, 2006), is applied to analyzing log or trace data in individual and collaborative learning settings to help understand students' temporal learning behaviors. As Paquette et al. (2021) noted, epistemic network analysis provides both statistical tests and networked visualizations for qualitative interpretations, overcoming some process mining limitations. Additionally, methods like constrained Sequential Pattern Discovery (cSPADE) (Kang et al., 2017; Ng et al., 2023; Wong et al., 2019; Zhichun Liu & Jewoong Moon, 2023), another form of sequence analysis, and the combination of process mining and clustering (Maldonado-Mahauad et al., 2018) provide innovative ways of capturing and analyzing the temporal and sequential characteristics of SRL.

However, another complication arises in choosing temporally focused analytical methods: deciding on the analytical direction—whether top-down or bottom-up—in which SRL skills and behaviors could be measured (Azevedo, 2014; Panadero et al., 2016). For instance, the sequential pattern mining approach (Zaki, 2001), being data-driven, and coherence analysis (Segedy et al., 2015), being theory-informed, provide unique insights into students' SRL-related behaviors. These two methods differentiate themselves in their fundamental analytical approaches. Sequential pattern mining is a bottom-up method that uncovers patterns directly from the observed data. On the other hand, coherence analysis exemplifies a top-down approach that leverages existing theoretical frameworks to conceptualize and interpret students' metacognitive behavior.

Sequential pattern mining is a data mining technique to uncover sequential patterns or event sequences in large databases (J. Han et al., 2022). This method analyzes frequent patterns of events to identify recurring patterns, such as transactions, time-stamped events, or activities. Unlike association rule mining, which focuses on co-occurring events, sequential pattern mining specifically targets the sequential relationship between events, emphasizing the temporal ordering and dependencies within a sequence. Sequential pattern mining also differs from lag-sequential analysis, which examines the strength and statistical relationships, such as transitional frequencies, between events at any given lag. Specifically, lag-sequential analysis focuses on calculating the probabilities of transitions between individual events or activities, making it effective for understanding the likelihood of one event following another in a sequence. In contrast, sequential pattern mining aims to identify frequent sequences of events within the entire dataset. It detects patterns that occur frequently, providing insights into common learning pathways and repeated behaviors within the dataset. There is, however, a great deal of overlap between the two methods, since the events within a sequential pattern are, by definition, in order and thus contain transitions. In our study, the primary interest lies in detecting frequent learning patterns across the entire dataset. Sequential pattern mining is well-suited for this purpose as it can uncover the most common sequences of learning activities, offering a broader view of learning behaviors.

Metacognitive Learning Strategies and Learning Patterns

Metacognitive strategies, a central component of SRL, encompass students' deliberate use of learning strategies to regulate their own learning process (Panadero, 2017). Identifying and understanding learning patterns associated with these strategies are crucial, since they can serve as valuable indicators of SRL usage, which can inform the design of AI-driven targeted interventions to improve students' SRL skills. Several studies have used sequential pattern mining to examine students' sequential learning patterns and behavior in computer-based online learning environments (S. Li et al., 2020; Mirriahi et al., 2016; Munshi et al., 2018; Shirvani Boroujeni & Dillenbourg, 2019; Siadaty et al., 2016; Zhang & Paquette, 2023). For instance, research in game-based learning environments has identified patterns in students' gameplay strategies or navigation sequences over time (Kang et al., 2017; Kang & Liu, 2022; Kinnebrew & Biswas, 2012; Rowe et al., 2015). Kang et al. (2017) and Kang and Liu (2022) utilized cSPADE to explore students' problem-solving behavior patterns within a serious game called Alien Rescue. The study focused on the behavior patterns of different performance groups and revealed distinct problem-solving strategies between high- and low-performing students.

In learning management systems, Poon et al. (2017) used sequential pattern mining to identify navigational patterns. Such pattern discovery in diverse learning environments assists in providing feedback to learners for a successful learning experience and offers insights for designers to enhance the learning environments (Perera et al., 2009). Regarding the application of sequential pattern mining in massive open online courses, Wong et al. (2019) utilized cSPADE to analyze log data, exploring differences in interaction patterns between students who viewed SRL prompt videos and those who did not. The findings indicated that SRL prompt viewers engaged with more course activities and exhibited a more consistent sequential pattern in completing them than their counterparts (Wong et al., 2019).

Building on this analysis of learning patterns, research further demonstrated how analyzing metacognitive strategies provides valuable insights into students' engagement in SRL (Segedy et al., 2015). For instance, coherence analysis (Segedy et al., 2015) provides a more theory-driven approach to understanding SRL compared to other learning analytics methods that are more data-driven. This approach measures metacognitive strategies during SRL by analyzing the coherence (i.e., how well two activities work together in sequence) of students' actions observed in online learning contexts. Focusing on coherence allows researchers to see beyond simple action and reaction, highlighting the importance of consistent, strategic behaviors in successful learning. The idea of measuring SRL skills via coherence analysis can be adapted to conceptualize numerous aspects of students' use of metacognitive strategies, tailored to specific learning settings and research contexts.

Numerous studies applied coherence analysis to assess students' employment of metacognitive strategies in online learning settings. For example, Bosch et al. (2021) examined the links among verbalized metacognition and learning, confusion, and metacognitive problem-solving strategies. Zhang et al. (2020) used coherence analysis in a computer-based learning environment called Betty's Brain to investigate the relationship between confusion and metacognitive strategies. Expanding upon their earlier work, Zhang et al. (2022) further utilized coherence analysis to explore the evolution of metacognitive strategy use, advancing the understanding of how metacognitive strategy use develops over time.

Study Participants and Research Context

Participants

We gathered behavioral data and survey responses from 210 college students who learned four different subtopics in statistics using a web-based learning environment. We used two sampling methods: in one sample, we recruited 112 students locally from a public research university in the Midwest region of the United States. Students who participated through this method received course credit upon completing the study. In the second sample, we recruited 98 students on Prolific, an online crowdsourcing platform that enables research with a diverse sample of students from U.S. colleges and universities (Peer et al., 2021). While Prolific allows researchers to filter potential participants based on various criteria, including demographic variables, we only restricted our selection to undergraduate students from either 2-year or 4-year community colleges or universities for eligibility in our study. The Prolific sample represented

62 unique colleges/universities, including 11 community colleges. We compensated each Prolific participant who completed our study with $15.

Ethics, Consent and Permissions

Before participating, students completed an IRB-approved consent form (IRB protocol #21019).

Demographics

We present students' self-reported demographic information to offer an insight into the diversity of our participants, even though not all demographic variables were examined in our analysis. Sample characteristics also serve to inform generalizability in meta-analytic research based on studies such as this one. Students self-described demographic characteristics, resulting in some fine-grained characteristics that had to be grouped together to protect privacy. Students' demographic information regarding race and ethnicity, gender, English as a first language, age, and class standing is described in the Appendix Tables 1-5.

Research Context

We developed a self-guided online learning system that allowed students to navigate educational content at their own pace. The study, spanning approximately 90 minutes, involved students engaging with the system to learn about introductory statistics. The participants began the study by completing a demographics survey. After completing the survey, participants took a pretest and were asked to guess their performance in a previous version of the same test without access to their actual scores. Following this, students engaged in a self-paced learning session for 60 minutes, during which their time remaining was displayed by a timer that operated exclusively during active software interaction to promote focus. The self-paced online learning environment included four distinct, illustratively presented subtopics with associated icons

(Figure 1). Each subtopic module included one reading, quiz, set of worked examples, and summary. Although students were not required to complete all the subtopics during the learning session, the platform allowed students to revisit and complete any activity multiple times, catering to their individual learning needs and preferences. As a final step in the study, participants took a posttest, which allowed us to measure learning gains by calculating the difference between their pretest and posttest scores.

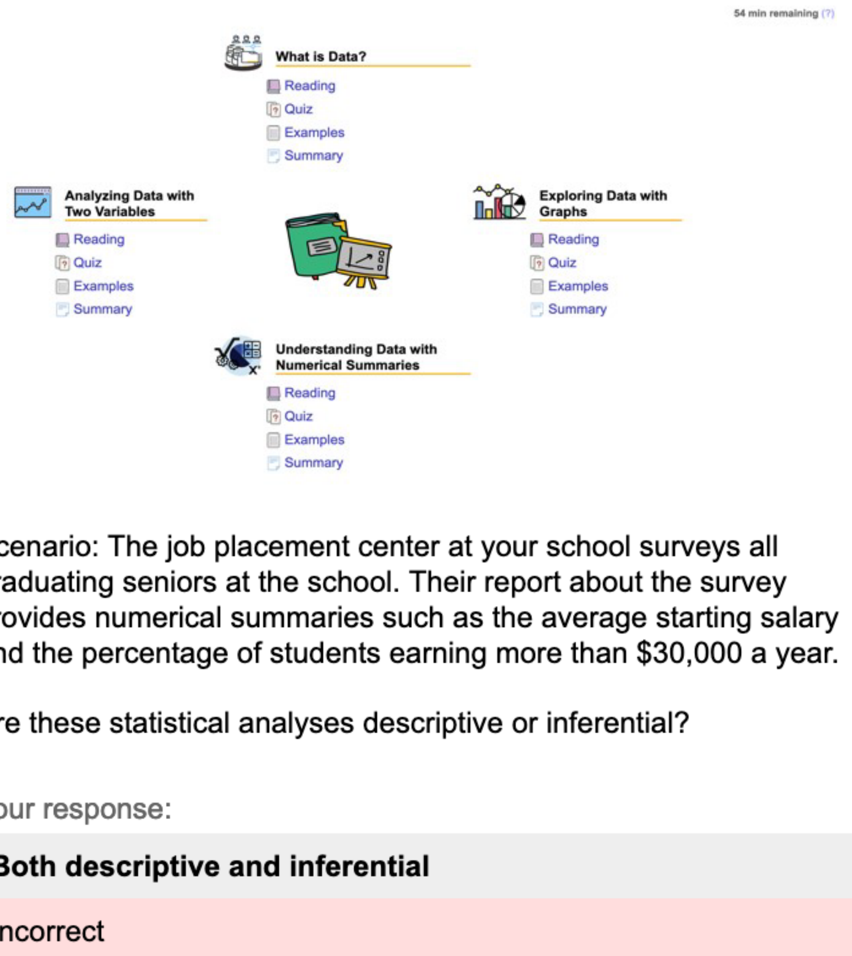
We developed pretest and posttest to evaluate knowledge on four subtopics covered in the learning material, with each test comprising 12 questions—three for each subtopic. In particular, pretests were designed to assess students' prior knowledge across 4 subtopics that were covered in the learning material. We calculated the correlation between students' actual pretest grade and their pretest score that they made immediately after taking the pretest ($r = .447$, $p < .001$), as verification that students' performance on the pretest aligns with their self-assessed understanding. Such alignment suggests that the pretest measures a construct that students are aware of, which could indirectly support its validity. From a convergent validity perspective, although self-assessed knowledge and actual knowledge differ, they are (ideally) closely related such that a positive correlation suggests the related constructs are indeed related. The correlation value of .447 does suggest a moderate positive association between students' actual pretest grades and their immediate guessing scores. This correlation serves as evidence that the pretest accurately reflects students' understanding of the concepts it is intended to measure, thereby aligning the pretest's objectives with students' perceptions of their own knowledge. Additionally, we report the correlation between pretest and posttest ($r = .480$, $p < .001$); although we would not expect the correlation to be perfect, since some students learn more than others, this correlation serves as evidence that pretest and posttest measure the same knowledge as intended.

Pretests and posttest were also designed to be as similar as possible in difficulty and subtopic coverage. To this end, we created two versions of the test, A and B, which were interchangeable as either pretests or posttests. To ensure similar difficulty levels for tests A and B, we calculated the percentage of students who answered each question correctly. This information is provided in the Appendix Table 6 along with the full questions from tests A and B. While most pairs of questions had similar correct response rates, indicating comparable difficulty levels for those specific questions in both tests, we noted some variations that suggest slight differences in difficulty. Despite these variations, tests A and B alternately featured questions with higher correct response rates. Therefore, while not all pairs of questions achieved exactly the same difficulty level, we ensured that tests A and B maintained a similar level of difficulty overall. Additionally, to minimize any ordering effects between tests A and B and ensure the reliability of our measurements, we employed a counterbalanced test order, where students were randomly assigned one of two versions: A or B. Students in version A began with test *A* as their pretest and test *B* for their posttest.

The test order was then reversed in version B to reduce effects due to test difficulty differences. Specifically, the random assignment of students to different test order neutralizes potential differences in test difficulty as it relates to statistical analysis of learning. Random assignment of test order evenly distributes potential differences in test difficulty across all students, thereby mitigating bias on average (i.e., in statistical point estimates of the mean) despite individual student-level biases in the learning gain estimate. Estimates of learning are therefore conservative, since the point estimate is unbiased but any potential test difficulty differences contribute variance to the estimate, decreasing statistical confidence. To establish content validity, we also asked four content experts (i.e., individuals with substantial postgraduate training in statistics) to match the randomly shuffled questions from tests A and B, which were interchangeably used as pretests and posttests, as described above. The experts were tasked with aligning each question from Test A to a corresponding question from Test B based on the statistical concepts they believed each question measured. All four experts achieved a correctness level of 100%, providing evidence that the questions from each test measure similar knowledge. We included full questions from Tests A and B in Appendix Table 6 to detail the structure and content of both the pretest and posttest.

Figure 1

Screenshot of a main menu of the learning software (top) and an illustration of an incorrect quiz question attempted by a student (bottom).



Note. If a student's quiz answer was wrong, we displayed that information, but did not display the correct answer.

Learning Activities and Subtopic Characteristics

The learning session comprised four distinct modules, each addressing a unique subtopic within introductory statistics. One subtopic, referred to as *Terminology* for concision, included comprehensive explanations and descriptions to enhance understanding of fundamental statistical concepts. It covered various related concepts, including descriptive and inferential statistics, the distinction between sample and population, the concept of margin of error, and its basic calculations. Furthermore, the subtopic contained the categorization of data types into categorical (including nominal and ordinal variables) and quantitative (comprising discrete and continuous variables). The *Graphs* subtopic focused on interpreting various graphs, particularly histograms. Emphasis was placed on comprehending histograms representing quantitative variables and identifying their distribution as unimodal, bimodal, or symmetric. The *Calculation* subtopic entailed computations of central measures, such as mean and mode, and dispersion measures, including variance and standard deviations. Finally, the *Amalgamation* subtopic covered various aspects like response and explanatory variables, confounding variables, and associations. Students also learned how to interpret scatterplots, understand correlation and correlation scatterplots, and grasp the properties of correlations. The Amalgamation subtopic necessitated a mix of competencies intrinsic to other subtopics, including the interpretation of graphs and the memorization of terminologies.

Each module comprised four distinct learning activities: reading, quizzes, examples, and summaries. Students had the flexibility to select the order of these activities, irrespective of the subtopic. Every activity served a distinct learning purpose. The reading activity, typically four to six pages per subtopic, provided comprehensive information about the subject matter. The quiz, consisting of around 10 questions, allowed students to assess their understanding of the material without any time limitations. Incorrect answers were flagged, but the correct answers were not revealed, promoting self-guided learning. The examples provided more than just correct answers to example questions; they demonstrated the proper problem-solving methods. Finally, the summary provided a concise recap of each module’s essential learning materials, allowing students to review each subtopic’s material quickly.

Analysis 1: Employing a Data-driven Approach

Data

We used 210 participants' trace data from the online learning system, which recorded their learning activities in real-time for our study. These trace data contained types of activities that students engaged in, activity durations, and test/quiz results recorded during the students' interactions with every stage of the software. In analysis 1, we leveraged sequential pattern mining and linear mixed-effect regressions. To implement the sequential pattern mining, we first transformed student's log data into a long format where every event was identified by a student ID, a learning activity (i.e., Read, Quiz, Example, and Summary), and an element ID (indicating the order of the learning activities). The element ID is crucial as it specifies the chronological order in which each learning activity occurred. For instance, in a learning sequence of Read → Quiz → Example, the "Quiz" learning activity would be assigned an element ID of 2 to indicate that it was the second activity in the sequence. Then, we transformed students' log data into two different long formats. The first, the overall learning activity list, contained the sequence of learning activities that students engaged in throughout the entire learning session, regardless of the subtopic. For the second long format data, subtopic-specific learning activity list, we extracted students' learning activities in relation to each subtopic (Terminology, Graphs, Calculation, and Amalgamation) individually. Therefore, we had four different subtopic-specific learning activity list for each subtopic. This transformation was necessary to enable us to investigate heterogeneity in learning patterns within SRL context across subtopics.

Analytic Framework and Methods

In this section, we provide a brief overview of our analytical approach, followed by detailed descriptions of each stage to help readers follow the methodological pipeline. The initial step involved using sequential pattern mining on students' overall learning activity sequence data to unveil commonly occurring learning patterns. We then identified and associated the frequently observed learning patterns potentially relevant to SRL-relevant strategies, as outlined in Table 1. This first step is detailed in the *Sequential Pattern Mining* subsection below. Once this groundwork was laid, we tallied each learning pattern's occurrences across subtopic-specific learning activity lists to examine the variations within the use of learning patterns within this SRL context across subtopics. Then, as a second step, we employed the frequencies of learning patterns potentially related to SRL-relevant strategies as variables within mixed-effects regression models (described in the *Analyzing Learning Patterns with Linear Mixed-Effects Regression* subsection) to address our research questions (RQ1–2).

Sequential Pattern Mining

We utilized the cSPADE algorithm (Zaki, 2000, 2001) for sequential pattern mining on our dataset. Specifically, we used the R package *arulesSequences* to implement cSPADE and discover frequent sequential learning patterns. cSPADE requires data in a long format and offers the flexibility to define parameters, such as the minimum support, which represents the threshold for the proportion of students utilizing a pattern for it to be considered frequent. Another constraint is the maximum gap, which sets the largest allowable time difference between elements in a sequence. Given that the appropriate settings for these parameters can vary based on the research context and objectives, many studies that leveraged cSPADE for discovering frequent patterns in the educational research domain determined these values based on the specifics of their research context (Kang et al., 2017; Kang & Liu, 2022; Ng et al., 2023; Wong et al., 2019; Zhichun Liu & Jewoong Moon, 2023). In our study, we used students' overall learning activity as a data input for cSPADE with minimum support value of 0.4 and a maximum gap value of 1. This minimum support value ensures that we only include sequences used by more than 40% of students in our results. The maximum gap value sets the largest allowable time difference between consecutive elements in a sequence. In our study, we defined a maximum gap value of 1, indicating that a sequence of two activities of interest should have at most one other activity between them. This particular constraint was chosen to align with our measurement of SRL constructs via coherence analysis for answering RQ2. In coherence analysis, we utilized a 5-minute window timeframe (as outlined in Analysis 2); for cSPADE, we observed that students, on average, spent approximately three minutes per learning activity, and thus there would typically be 2–3 activities overlapping with any given 5 minutes (i.e., a maximum gap of 1), approximately matching the timeframe used for coherence analysis.

After applying cSPADE to the overall learning activity list, we obtained the most common sequences of learning activity patterns that students engaged in and the corresponding support values, which indicates the proportion of students who engaged in each frequent learning pattern at least once. For instance, an example of a frequent learning pattern could be *Read → Quiz* with support value .75, which implies that 75% of the students engaged in a reading activity then a quiz activity at least once throughout the entire learning session.

Associating Frequent Learning Patterns to Potential SRL-relevant Strategies

We examined frequent learning sequences to relate these recurring learning patterns to potential SRL-relevant strategies that are theoretically grounded in the literature (Corrin et al., 2017; Sonnenberg & Bannert, 2015; Zimmerman & Pons, 1986). We mainly adopted Zimmerman's 14 classes of SRL strategies (Zimmerman & Pons, 1986) as a framework to relate each frequent learning pattern to a potential SRL-relevant strategy. Zimmerman and Pons developed these 14 types of SRL strategies to assess students' application of SRL in naturalistic settings. In their SRL strategy schema, they defined SRL strategy as actions directed at acquiring information or skills, such that the actions involve agency, purpose (goals), and instrumentality self-perceptions by a learner (Zimmerman & Pons, 1986). Zimmerman and Pons's SRL strategies focus on evaluating students' active SRL behaviors in terms of their actions. While Zimmerman's three-phase SRL model (Zimmerman & Moylan, 2009) describes SRL through distinct phases such as forethought, performance, and self-reflection, the classification of 14 SRL strategies delves deeper into evaluating students' active application of these strategies. Notably, these strategies, particularly those centered on action, may align closely with the performance phase of Zimmerman's model, where learners are actively employing SRL strategies and behaviors, as opposed to phases before learning (i.e., forethought) or after (i.e., reflection).

Given our focus on investigating students' potential use of SRL-relevant strategies based on their learning patterns during their active learning phase using trace data from computer interactions, Zimmerman's SRL strategy classifications serve as a fitting framework for our study—provided that we operationalize the potential use of SRL-relevant strategies in terms of behaviors that are possible in our computer-based learning context. SRL encompasses not just the observable use of SRL strategies by students, but also their motivational aspects and self-perceptions, which are inherently internal and often difficult to measure solely through trace data extracted from online learning platforms. However, capturing both of these aspects, especially students "agency, purpose, or instrumentality self-perceptions" as described by Zimmerman and Pons, is challenging when only trace data is available. The feasibility of collecting self-reported data on SRL varies, making it essential, as we aim in this study, to devise methods to measure and conceptualize SRL solely through the analysis of trace data. In this study, we adopt a process-based perspective to understand SRL, focusing on investigating students' observable actions as inferred from trace data.

Therefore, we note that our approach to conceptualizing the potential usage of SRL-relevant strategies does not encompass students' "agency, purpose, or instrumentality self-perceptions," as directly measured by self-reported surveys, which are subjective to student's own beliefs and perceptions. However, our SRL-relevant strategy conceptualization does aim to capture students' agency, purpose, or instrumentality to the extent that they are apparent through trace data, rather than through self-reported data. Although trace data cannot capture all aspects of SRL, at least not equally, it can still provide valuable insights into students' agency and purpose. For instance, frequent and purposeful engagement with specific learning strategies can indicate a high level of agency and goal orientation. Learning patterns such as regular revisiting of reading or repeating quiz takings can reflect a student's purpose and strategic approach towards achieving their goals.

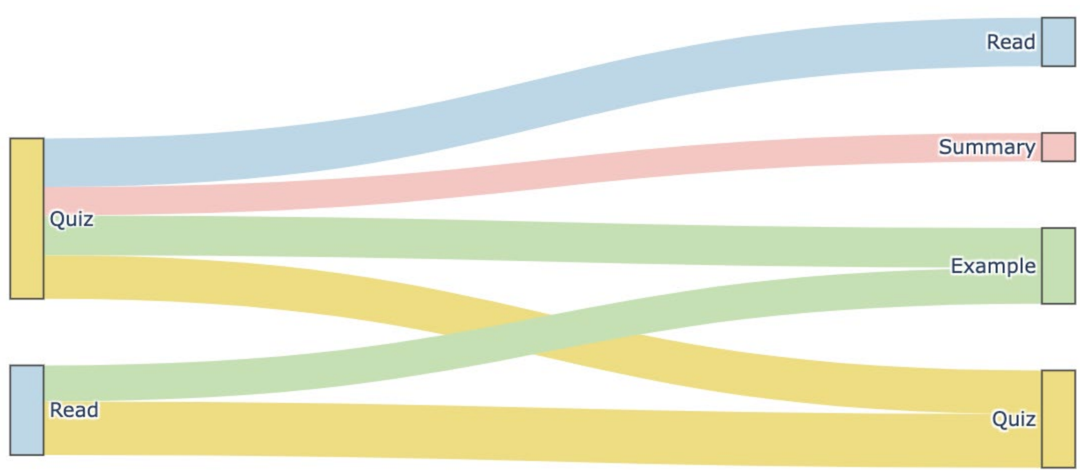
Table 1 details each frequent learning pattern, its support value, and potentially associated SRL-relevant strategies. For instance, the most prevalent learning pattern identified was the *Read → Quiz*, which could potentially imply the use of seeking information SRL-relevant strategy. Further, we provide a detailed description in Table 1 on how each learning pattern is potentially associated with SRL-relevant strategies. We highlight that the frequent learning pattern may potentially imply student's use of SRL-relevant strategies, thus does not strictly indicate student's use of SRL strategies. For instance, the *Read → Quiz* sequence could potentially imply students' use of *seeking evaluation* strategy (Zimmerman & Pons, 1986). When learners read material and then take a quiz, they are assessing their comprehension and recall of the content. By taking the quiz, students can evaluate the quality or progress of their understanding based on their performance. Learning patterns such as *Quiz → Read* and *Quiz → Summary,* where a student takes the quiz and goes on to reading or summary, could potentially be associated with student's use of *keeping records and monitoring* (Zimmerman & Pons, 1986), *seeking information* (Zimmerman & Pons, 1986), and *search* (Sonnenberg & Bannert, 2015). Students engaging in these learning patterns are likely to seek information relevant to their quiz attempts to enhance their understanding. Further, these learning patterns could potentially indicate that students are aware of knowledge gaps found by taking quizzes and actively search for specific material to address these gaps.

We potentially related *Quiz → Quiz* learning pattern with *rehearsing and memorizing* (Zimmerman & Pons, 1986) and *repeating* (Sonnenberg & Bannert, 2015) SRL-relevant strategies. Engaging in continuous self-assessments allows students to rehearse and help them identify errors and knowledge gaps. Moreover, the act of retaking quizzes aligns with the SRL-relevant strategy of repeating, as it provides continuous practice and aids in the deepening of understanding. *Quiz → Examples* learning pattern could potentially indicate the use of *help-seeking* (Corrin et al., 2017), *keeping records and monitoring* (Zimmerman & Pons, 1986), and *seeking information* (Zimmerman & Pons, 1986) SRL-relevant strategies. This learning pattern can possibly imply students' proactive efforts to clarify doubts by seeking help through reviewing example questions like those in the quiz. By referring to examples after quizzes, students monitor their performance and track progress, ensuring they comprehend the material. We associated *Read → Examples* potentially with *elaboration* (Sonnenberg & Bannert, 2015) and *seeking information* (Zimmerman & Pons, 1986) SRL-relevant strategies. *Elaboration* involves deeper processing through activities such as paraphrasing, connecting, and inferring (Sonnenberg & Bannert, 2015). This learning pattern possibly suggests that students engage in detailed examination and integration of the material by connecting reading content with practical examples. Additionally, by going over the examples, students actively seek relevant information to enhance their understanding.

Visualizing frequent learning patterns among students can be a challenging task, especially given the variability in sequence lengths and the number of elements within sequences depending on the research objectives. We used a Sankey diagram (Figure 2) to illustrate the frequent learning patterns potentially linked to specific SRL-relevant strategies as outlined in Table 1. The diagrams' links effectively display the sequence in which each learning pattern is employed within SRL context. The width of each link signifies its support level; broader links indicate a higher number of students engaging in a specific learning pattern.

Figure 2

Sankey Diagram Displaying Students' Usage of Frequent Learning Patterns



Note. To interpret the diagram, begin from the leftmost label, *Quiz*. This starting point branches into four distinct frequent learning patterns: transitioning from *Quiz* to *Read*, *Quiz* to *Summary*, *Quiz* to *Example*, and *Quiz* to another *Quiz*.

Table 1

Proposed Alignment Between SRL-relevant Strategy and the Frequent Learning Patterns Found via the cSPADE Algorithm

Frequent Learning Pattern	Support	SRL-relevant strategy (specific subcategory)	Description
Read → Quiz	0.85	- *Seeking evaluation* (Zimmerman & Pons, 1986)	When students read material and then take a quiz on it, they are evaluating their understanding and recall of the material they just read. After taking the quiz, they can gauge the quality or progress of their work based on their performance.
Quiz → Read	0.77	- *Keeping records and monitoring* (Zimmerman & Pons, 1986) - *Seeking information* (Zimmerman & Pons, 1986) - *Search* (Sonnenberg & Bannert, 2015)	Students taking the quiz and then reading the main material signifies that students are aware of the knowledge gap and might specifically look for information to address the gaps.
Quiz → Quiz	0.69	- *Rehearsing and memorizing* (Zimmerman & Pons, 1986) - *Repeating* (Sonnenberg & Bannert, 2015)	When students encounter the first quiz, they are prompted to recall specific information. By the second quiz, they are not just accessing their foundational understanding but also relying on memory from the previous quiz attempt.
Quiz → Examples	0.63	- *Keeping records and monitoring* (Zimmerman & Pons, 1986) - *Seeking information* (Zimmerman & Pons, 1986) - *Help-seeking* (Corrin, de Barba, & Bakharia, 2017)	After taking the quiz, students are trying to make an effort to gather specific, detailed information on how to approach or solve problems correctly.
Read → Examples	0.57	- *Seeking information* (Zimmerman & Pons, 1986) - *Elaboration* (Sonnenberg & Bannert, 2015)	Illustrates student-initiated efforts to seek additional knowledge from additional resources to bolster their learning. Students are actively seeking clarity and deeper understanding as Examples provides them with detailed worked-out problems with explanations on how to approach solving the problem.
Quiz → Summary	0.45	- *Keeping records and monitoring* (Zimmerman & Pons, 1986)	Students are not only keeping records of their quiz performance but are also actively seeking to enhance their understanding through the

		- *Seeking information* (Zimmerman & Pons, 1986) - *Search* (Sonnenberg & Bannert, 2015)	supplementary information provided in the summary. This dual approach allows them to both identify areas of improvement from their quiz results and address those areas by going through a summary.

Note. The left most column lists the frequent learning patterns, followed by a column indicating their support values, which reflect the prevalence of these patterns among the students. The next columns detail the potential association with corresponding SRL-relevant strategies found in the literature and provide descriptions of these possible associations.

Analyzing Learning Patterns with Linear Mixed-Effects Regression

Using students' subtopic-specific learning activity lists, we counted the occurrence of each frequent learning pattern for each subtopic (Table 1). By examining the frequencies of learning patterns across subtopics, we determined whether students adjusted their learning patterns with varying frequency across subtopics. We then employed these frequencies of subtopic-specific learning patterns as variables in a linear mixed-effects regression. However, as discussed in the previous section, one of the limitations of using cSPADE is that the results (i.e., the frequent learning patterns) from cSPADE do not afford an inferential, statistical interpretation. In this study, we overcame this limitation by arranging the cSPADE results such that they are suitable for follow-up linear mixed-effects regression modeling. All models included subtopic-wise frequency of learning patterns and learning gain as dependent variables. For all models, we checked the assumptions of linear regression (linear relationship, independence, homoscedasticity, and normality). All regression models included a random intercept for participant ID to account for the hierarchical nature of the data (i.e., many observed behaviors per student). Such an approach allowed us to consider individual differences at the baseline level. We report standardized betas as effect sizes in situations where predictors and outcomes were continuous, or partially standardized betas where predictors were categorical (e.g., subtopic ID).

For RQ1, we analyzed the occurrences of learning patterns as dependent variables and subtopic names (treated as factor variables) as independent variables. For RQ2, we also modeled each learning pattern frequency as the dependent variable in a model, but included different predictor variables: type of subtopics, learning gain, and prior knowledge (as measured by pretest score). We note that our focus is to investigate whether we observe variations in the frequencies of learning patterns within an SRL context across different subtopics. Therefore, to address our research question, we used learning patterns as dependent variables, rather than predictors. Learning gain was measured by taking the difference between students' posttest and pretest grades to assess how much students improved in their understanding of the study material throughout the learning session. We also considered interactions between these predictors. Specifically, we hypothesize that the influence of prior knowledge and learning gain on the count of learning patterns might vary depending on the study subtopic. Thus, we included interaction terms between the study subtopic and the other three predictors. These terms allowed us to assess whether the heterogeneous effects of pretest score and learning gain on the occurrence of learning patterns are explained by the specific subtopic under study. We note that in RQ2, we are interested in investigating whether students' prior knowledge, learning gain, and their interactions across subtopics have predictive power regarding engagement with learning patterns in an SRL context. Therefore, learning gain and prior knowledge are used as independent variables, allowing us to explore how variations in these factors are associated with using learning patterns.

Analysis 1 Results

Before answering RQ1 and RQ2, we present an overview of learning patterns along with potentially associated SRL-relevant strategy usage across subtopics to examine whether there was a varied distribution of these learning patterns with possibly related SRL-relevant strategies among students within different subtopics (Table 2). Specifically, the frequencies indicate the proportion of students who engaged in each learning pattern at least once, across different subtopics. Across all subtopics, we found that students predominantly engaged in the learning pattern of *Read → Quiz* which we propose is associated with the SRL-relevant strategy of *seeking evaluation* (as shown in Table 1). However, the extent to which students engaged in this learning pattern varied notably across different subtopics. For instance, engagement rates were observed to be 79.4% for Terminology, 65.1% for Graphs, 58.1% for Calculation, and 66.8% for Amalgamation. Further, our analysis revealed that certain learning patterns exhibited higher frequencies in specific subtopics compared to others. For instance, the learning pattern of *Quiz → Quiz*, which we propose is associated with the use of *rehearsing and memorizing* and *repeating* SRL-relevant strategies was found to be a common learning pattern within the Graphs subtopic (34.0%). In contrast, the *Read → Example* learning pattern, which we propose is associated with the *seeking information* SRL-relevant strategy, was particularly prominent within the Calculation subtopic, accounting for 31.0% of the learning patterns students engaged in. The observed subtopic-specific learning patterns possibly associated with SRL-relevant strategies hint at the possibility that students may adapt their choice of learning patterns based on the subtopic they are studying. For instance, the prevalence of the learning pattern, *Quiz → Quiz,* which we propose is associated with the use of *rehearsing and memorizing* and *repeating* SRL-relevant strategies in the Graphs subtopic might suggest that recalling specific information during the initial quiz, and then reinforcing that memory in subsequent quizzes are particularly beneficial for understanding graphical data. Meanwhile, the high occurrence of the *Read → Quiz* learning pattern, possibly implying the use of *seeking information* and *elaboration* SRL-relevant strategies, indicates that students may benefit from seeking detailed examples after their readings to further consolidate their understanding, especially when dealing with computational or problem-solving tasks.

We conducted an additional analysis to investigate the mindfulness of students' engagement in these learning patterns in SRL context. To this end, we randomly shuffled the order of learning activities that students engaged in for each subtopic using the second long format data, which is described in the *Data* subsection of the *Analysis 1: Employing a Data-driven Approach* section. Specifically, we shuffled each subtopic's learning activity list using the "shuffle" function in Python. Shuffling each subtopic's learning activity list simulates a null distribution in which students engage in the observed activities with no intentionality, i.e., without selecting their next activity based on previous activities. We adopted the same approach for calculating the frequencies of learning patterns for each subtopic, as provided in Table 2, when calculating the frequencies of each learning pattern using the shuffled data. By comparing the frequencies of learning patterns in the original data with those in the shuffled data, we aimed to investigate whether students' engagement in learning patterns was intentional or merely random.

Differing frequencies between randomly shuffled and original data in learning sequences would provide insight into the intentionality behind students' actions. The rationale behind this comparison is that intentional actions tend to produce consistent learning patterns that are unlikely to occur by chance. For instance, if students are consciously engaging in learning patterns (e.g., intentionally following a *Read* → *Quiz* sequence to check their understanding), we would expect the frequency of such learning pattern in the original data to differ from it in the shuffled data, where the order of actions is randomized. This difference suggests that students are deliberately choosing to engage in a specific learning pattern. In contrast, if students are engaging in learning patterns without clear intention—perhaps clicking on activities without much thought—we would expect the frequencies in the original and shuffled data to be similar. This similarity would indicate that the sequences are not the result of intentionally selecting an action based on the previous action(s).

This approach to discerning whether students' learning pattern is happening by a random chance or not is taken from research on sequence mining and permutation tests, which are used to identify patterns that significantly deviate from what would occur by chance (Pinxteren & Calders, 2021; Tonon & Vandin, 2019; Zhang et al., 2024). Specifically, Zhang et al. (2024) demonstrated the effectiveness of permutation tests in identifying statistically significant and nonredundant patterns in educational data. The permutation test, as described by Zhang et al. (2024), is directly analogous to our approach since it involves creating a baseline of random data against which the original data is compared. In their study, permutation tests are used to shuffle the sequence of events in educational data to determine which patterns occur more frequently than would be expected by chance. This approach filters out random patterns, highlighting those that are statistically significant and likely to represent intentional patterns.

Our analysis revealed differing frequencies for learning patterns across subtopics. The frequency of the learning sequence (*Read* → *Quiz*) was 79.4% in the original data but dropped to 34.8% in the randomly shuffled data. This difference in frequencies implies that students were more likely to engage in the *Read* → *Quiz* sequence intentionally rather than randomly.

Similarly, decreased frequencies for the *Read → Quiz* sequence across all subtopics in the shuffled data further support the idea that students' engagement in this sequence was deliberate and less mindless. Furthermore, we observed varying frequencies in all other learning patterns as well. For instance, for certain specific learning pattern such as *Quiz → Read*, we observed increased frequencies in the shuffled data, suggesting that students engaged in these learning patterns less frequently than expected by random chance in the original data. Moreover, we observed similar frequencies in a few cases of the learning patterns across both the original and shuffled data, suggesting that certain sequences might be less dependent on intentionality. Although we acknowledge that some students might have engaged in these learning patterns mindlessly, the consistent differences in frequencies between the original and shuffled data suggest that a substantial portion of students were engaging in these learning patterns intentionally.

Table 2

Frequencies of Learning Patterns across Subtopics

Sequence	Terminology	Terminology (shuffled)	Graph	Graph (shuffled)	Calculation	Calculation (shuffled)	Amalgamation	Amalgamation (shuffled)
Read → Quiz	79.4%	34.8%	65.1%	34.0%	58.1%	32.5%	66.8%	28.8%
Quiz → Read	17.1%	32.4%	23.4%	37.3%	22.7%	30.0%	15.4%	28.4%
Quiz → Quiz	28.1%	32.0%	34.0%	33.0%	25.1%	26.6%	21.2%	22.1%
Quiz → Examples	47.6%	24.8%	33.0%	25.8%	23.3%	20.2%	32.7%	21.1%
Read → Examples	25.2%	23.8%	23.4%	21.5%	31.0%	18.7%	23.6%	23.1%
Quiz → Summary	27.1%	35.2%	24.4%	28.3%	24.1%	25.6%	22.1%	23.6%

Note. The percentages do not add up to 100% for each subtopic because the frequencies indicate the proportion of students engaged in each learning pattern across subtopics. The corresponding column labeled "shuffled" for each subtopic indicates the frequency of learning pattern calculated using shuffled data.

RQ1: Relationship Between Frequency of SRL-relevant Strategies and Subtopic

Across different subtopics, we found heterogeneous frequencies of learning patterns potentially associated with SRL-relevant strategies such as *Read → Quiz, Quiz → Quiz, and Quiz → Summary*, depending on the subtopic of the study. The *Read → Quiz* learning pattern, possibly linked with the SRL-relevant strategy of seeking information, was significantly less prevalent in the Graphs subtopic (β = -.397, 95% *CI* [-.549, -.245], $p < .001$), the Calculation subtopic (β = -.492, 95% *CI* [-.644, -.341], $p < .001$), and the Amalgamation subtopic (β = -.405, 95% *CI* [-.557, -.253], $p < .001$) compared to the Terminology subtopic. A comprehensive regression table is provided in Appendix Table 7. On the other hand, the *Quiz → Quiz* learning pattern, potentially implying the use of *rehearsing and memorizing* and *repeating* SRL-relevant strategies, was more frequent in both the Terminology and Graphs subtopics compared to Calculation and Amalgamation. This was evident for Terminology (β = .167, 95% *CI* [.012, .229], $p = .030$) and Graphs (β = .287, 95% *CI* [.099, .316], $p < .001$) compared to Calculation. Additionally, the frequency of the learning pattern, *Quiz → Summary,* potentially associated with *keeping records and monitoring*, *seeking information*, and *search* SRL-relevant strategies, was significantly higher in the Terminology subtopic compared to Graph (b = .29, 95% *CI* [.134, .447], $p < .001$), Calculation (b = .489, 95% *CI* [.332, .645], $p < .001$), and Amalgamation (b = 0.36, 95% *CI* [.203, .516], $p < .001$). The significant variability we observed in occurrences of learning patterns across subtopics suggests that students might adapt their frequency of specific learning patterns based on the subtopic under study.

RQ2: Subtopic Heterogeneity in the Relationship Between SRL-relevant Strategies, Learning Gain, and Prior Knowledge

We found significant results regarding the relationships between learning gain and the frequency of various learning patterns in an SRL context. In particular, we found that for the Graphs subtopic (relative to the reference level, Terminology), learning gain was significantly negatively related to the frequency of the learning pattern *Quiz* → *Quiz*, which we propose implies the use of *rehearsing and memorizing* and *repeating* SRL-relevant strategies (β = -.234, 95% *CI* [-.445, -.023], p = .031*)*. A comprehensive regression table is provided in Appendix Table 8. This result implies that—compared to studying the Terminology subtopic—when students study the Calculation subtopic, as their learning gain increases, the frequency of engaging in the *Quiz* → *Quiz* learning pattern decreases.

Similarly, compared to the Graphs subtopic, students' learning gains were significantly negatively associated to the *Quiz* → *Quiz* learning pattern in Calculation (β = -.352, 95% *CI* [-.555, -.148], p < .001). On the other hand, compared to the Calculation subtopic, learning gain was significantly positively related to the frequency of engaging in the *Quiz* → *Quiz* learning pattern in Amalgamation (b = .225, 95% *CI* [.011, .439], p = .225). For the *Quiz* → *Example* learning pattern, which we propose is associated with the SRL-relevant strategies of *keeping records and monitoring*, *seeking information*, and *help-seeking*, we found that, similar to the *Quiz* → *Quiz* learning pattern, learning gain was significantly negatively associated with the frequency of engaging in the *Quiz* → *Example* learning pattern for the Calculation subtopic compared to the Terminology subtopic (b = -.289, 95% CI [-.527, -.051], p = .018). These varying relationships between learning gains and prior knowledge, and the use of learning patterns within this SRL context suggest that students with different levels of prior knowledge and learning gain might adjust their use of learning patterns depending on the subtopics.

Analysis 2: Employing a Theory-driven Approach: Coherence Analysis

In the second analysis, we employed coherence analysis and linear mixed-effect regressions to examine the relationship between SRL measures and learning gains across subtopics (RQ3).

Method

Coherence Analysis

Coherence Analysis (CA) is a theory-based method that measures the use of metacognitive strategies via the order and timing of learning activities (Segedy et al., 2015). CA quantifies the extent to which specific learning activities work together (i.e., are coherent) to enact certain metacognitive strategies. For instance, if a student takes a quiz, then reviews the material relevant to any incorrect quiz answers, both taking the quiz and revisiting the content (such as reading) exemplify coherence. Coherent actions implicitly signify the utilization of metacognitive strategies, given that such an action involves a student assessing information gleaned from previous activities (such as perusing relevant material) and modulating their current actions (like taking a quiz) based on this information (Segedy et al., 2015; Zhang et al., 2020). Coherent actions need not be sequential, but it is necessary to constrain the time interval between those actions since it is less clear (and perhaps less likely) that one action is informed by the results of a specific previous action if that previous action is in the distant past. Previous research in the context of Betty's Brain revealed that students typically utilized the information they encountered within five minutes of encountering it: coherent actions within this time span were positively correlated with assessment scores within a learning session as well as learning gains across a whole session (Segedy et al., 2015). We developed CA measures based on metacognitive theory to capture students' use of metacognitive regulation during active learning, focusing on SRL skills like planning, monitoring, and managing their use of skills (Veenman, 2016).

We defined two universal CA measures—coherent quiz and coherent reading—and computed these two CA measures for each subtopic, resulting in four unique, subtopic-specific sets of CA measurements. These subtopic-specific CA measurements capture the variability in SRL-relevant strategies as students navigate each subtopic. The "coherent quiz" CA measure refers to the cumulative time a student spent engaging in reading activities within the five minutes prior to taking quizzes on the topics in those readings. The reading activities encompassed three types: studying the primary reading material, reading worked-out examples, and reading summary pages. Collectively, these three actions are referred to as "reading" actions within the context of this study. As such, the coherent quiz action was quantified based on the total time students allocated to reading activities before undertaking the quiz within a five-minute window. Coherent quiz behavior indicates that students are thoughtfully allocating their time to read and understand the necessary information before testing their understanding of that information by taking the quiz. Coherent quiz behavior thus exemplifies one usage of metacognitive strategies as students self-regulate their review and assessment processes. Similarly, the decision to utilize the information within a specific time frame (the five-minute window) indicates the students' awareness of the relevance and retention of the information.

A related CA measure, “coherent reading” refers to students’ time spent reading material related to the questions they missed in the quiz. We calculated coherent reading by tallying the time students spent studying the related material of missed quiz questions within a five-minute window following the quiz. Coherent reading represents a metacognitive strategy that comes after a quiz, wherein students identified their knowledge gaps through quiz results and immediately dedicated time to addressing these gaps by focusing on the specific areas of misunderstanding. Such an approach highlights a student’s capability to monitor their learning progress, recognize their errors, and take the necessary action to improve—key elements of metacognitive strategy use. Hence, coherent reading serves as a valuable indicator of the application of metacognitive strategies in the learning process.

The effectiveness of CA constructs was demonstrated by Segedy et al. (2015) in the Betty’s Brain learning platform, where students are expected to teach a virtual agent by developing a causal map. The researchers measured five CA constructs: edit frequency, unsupported edit percentage, information viewing time, potential generation time, and used potential time (Segedy et al., 2015). Potential generation and information viewing time are closely aligned with coherent quiz and coherent reading constructs from this study. Potential generation is quantified by the amount of time students spend viewing information that could support their subsequent action, which is editing causal map in Betty’s Brain. Likewise, coherent quiz is measured by the total time students spend reading relevant material before taking the quiz, where reading time supports students’ quiz attempt. Similarly, information viewing time refers to the time spent reviewing graded answers or resource pages, corresponding to coherent reading measurement, which totals the time spent reviewing material related to missed quiz questions.

Investigating the Relationship Between Metacognitive Strategy Use and Learning Gain with Linear Mixed-effects Regression

We explored the relationships between the topic, coherent reading, and coherent quiz actions on learning gains via mixed-effects regression. In RQ3, learning gain was the dependent variable, and predictor variables were the study subtopic name and two different CA measures: coherent reading and coherent quiz. We also included the interaction terms between subtopics and CA measures to examine how the association between learning gain and coherent behaviors differed across different subtopics. The regression model included a random effect for participant ID to account for the hierarchical nature of the data.

Analysis 2 Results

RQ3: Relationship Between the Use of Metacognitive Strategies and Learning Gain across Subtopics

For the main effects of coherent reading and coherent quiz, we observed a statistically significant negative effect of coherent quiz on learning gain only when the Calculation subtopic was the reference variable (b = -4.393, 95% CI [-8.172, -.610], p = .023). A comprehensive regression table is provided in Appendix Table 9. We found a significant negative interaction between the Calculation subtopic and coherent quiz measures, compared to the Terminology subtopic (b = -6.212, 95% CI [-12.043, -.373], p = .037) and compared to the Graphs subtopic (b = -11.066, 95% CI [-19.685, -2.441], p = .013). These negative interactions indicate that coherent quiz behavior was less effective for learning in the Calculation subtopic, relative to others. As students spent more time engaging in coherent quiz actions, the learning gain decreased on that subtopic compared to other subtopics. One potential explanation is that the distinct nature of the Calculation subtopic, which heavily focuses heavily on solving mathematical problems, may not benefit as much from preparatory reading before a quiz, as is demonstrated by a coherent quiz activity. This lack of benefit might be because direct engagement with quiz questions that require problem-solving might be more advantageous for enhancing understanding of the material within the Calculation subtopic. In contrast, for subtopics such as Graphs or Terminology, where conceptual understanding is crucial, spending time on reading before a quiz might improve students' comprehension, although significant results were not observed. This divergence in SRL-relevant strategy effectiveness highlights the need for a more context-dependent approach to supporting students' SRL in computer-based online learning environments. Further, these findings imply that certain SRL behaviors may be more beneficial than others within specific subtopics, indicating that the effectiveness of SRL-relevant strategies might vary, even within a single domain.

Discussion

By leveraging data collected from 210 college students engaged in a computer-based learning environment for introductory statistics with diverse subtopics, we addressed three research questions. At a high level, we found:

RQ1. We observed a significant variability in frequencies of learning patterns across subtopics within an SRL context.

RQ2. Students with different levels of prior knowledge and learning gains exhibited varying degrees of engagement of learning patterns potentially associated with SRL-relevant strategies across subtopics.

RQ3. In the calculation subtopic, engaging in coherent quiz activities had a negative impact on learning gains.

Theoretical Implications

Our findings contribute to the refinement of both SRL theory and its practical application within computer-based learning environments. We first situate findings regarding learning patterns potentially associated with SRL-relevant strategies and outcome heterogeneity within SRL models, contributing to understanding and refining these models. Students employing different learning patterns across various subtopics within an SRL context (RQ1) suggests that the dimension of contextual variability could be incorporated into SRL models to account for variations of learning patterns within an SRL context depending on the specific subtopic of the study. In Zimmerman's model and in Winne and Hadwin's model, we suggest adding the dimension of contextual variability, which would acknowledge and account for variations of learning patterns possibly related to SRL-relevant strategies depending on the specific subtopic of the study. Specifically, the *performance* phase of Zimmerman's cyclical model could be augmented to reflect that students might execute the metacognitive strategies differently across subtopics, and that doing so is beneficial for learning when variations in strategy are congruent with heterogeneous task demands. Similarly, in the *operation* phase of Winne and Hadwin's model, where students employ learning strategies, we propose an enhanced emphasis on the role of subtopic heterogeneity. This involves students not just deploying a generic set of strategies across all tasks but adaptively selecting and modifying their strategies based on each subtopic's specific demands and nature. By doing so, students can better align their efforts with the unique requirements of different subtopics, thereby potentially enhancing their overall learning effectiveness. We investigated heterogeneity in learning patterns within an SRL context utilizing complementary data-driven and theory-informed methods, which also could provide opportunities to extend SRL theory by considering how the findings of open-ended, data-driven methods indicate areas for expanding theory-driven methods (i.e., CA, in this case). Specifically, we demonstrated how using a data-driven method, specifically sequential pattern mining, helped us in potentially associating students' frequent learning patterns to SRL-relevant strategies. This possible alignment, combined with linear regression, allowed us to uncover the potentially heterogeneous nature of learning patterns within an SRL context. Furthermore, these preliminary insights into heterogeneity prompted us to examine it by constructing CA measures which were designed to capture students' use of metacognitive regulation. However, since this theory-driven approach can be adapted to measure other aspects of students' learning strategies, our study also highlights how the theory-driven coherence analysis, in this case, can be further expanded to explore other manifestations of SRL, thereby extending SRL theory.

Practical Implications

Our findings suggest that it becomes crucial to consider the heterogeneous nature of learning patterns potentially associated with SRL-relevant strategies and outcomes when it comes to designing SRL-supportive learning environments, given that refinements to SRL theoretical models should result in corresponding changes to the ways that SRL skills that are taught to students (e.g., to set expectations for the outcome of applying a particular SRL skill in context). The variability in learning patterns across different subtopics in an SRL context (RQ1-2), combined with the finding that the effectiveness of metacognitive strategies is not uniform across all subtopics (RQ3), collectively offers insights for developing personalized SRL supporting tools in computer-based learning environments.

These insights direct us towards the development of SRL supporting tools that are not merely adaptive to students, but also to specific content they are engaging with. Such tools would benefit from incorporating more tailored approaches that can do more than track and encourage frequent SRL-relevant strategies; they would also analyze the effectiveness of these strategies in relation to the learner's performance in the current context. By doing so, SRL supporting tools can guide learners away from over-relying on strategies that are less effective for a given subtopic and steer them towards alternative approaches that are better suited to the demands of the subtopics. Furthermore, the potential of data-driven AI systems to develop such personalized SRL supporting tools is significant. By data-driven AI systems, we refer to artificial intelligent systems that leverage data-driven algorithms or advanced computational techniques such as machine learning, explainable AI methods, or predictive analytics. Specifically, by leveraging insights into the variability of learning patterns effectiveness across subtopics and individual differences in an SRL context, AI systems can use machine learning algorithms to predict or identify the most effective SRL-relevant strategies for a given subtopic, and subsequently provide personalized recommendations. For instance, adopting a personalized and context-sensitive approach would enable the SRL supporting systems to recommend more effective strategies for calculation-related subtopics, where attempting repeated quizzes may be effective in computer-based learning environments. However, it still necessitates further investigation into the heterogeneous characteristics of SRL and its effectiveness on learning in future research.

Methodological Implications

In addition to the theoretical and practical contributions, our work introduces a complementary methodological approach that integrates cSPADE and mixed-effects regression modeling. This method enhances the utility of cSPADE's sequential pattern outputs by enabling statistically robust explorations of temporal patterns in SRL. Although SRL measurement is a complex and evolving field (Fan et al., 2022; Hilpert et al., 2023), as our understanding deepens and our analytical tools improve, we anticipate the emergence of new measurement approaches that offer even richer insights into SRL. This study is a step in that direction, demonstrating the power of combining complementary analytical approaches in uncovering the complex dynamics of SRL. Our approach paves the way for future research and practical implementations that consider the multidimensional nature of heterogeneous subtopics and its association with students' learning gain and prior knowledge in both theory and application.

Limitations

In this section, we discuss several limitations of our study. First, the online learning environment we developed might differ from some other online learning environments, such as semester-long computer-based courses where the duration of the learning period is longer and students typically have more diverse options for learning activities. Further research is needed to explore whether our findings regarding heterogeneity in learning patterns within an SRL context can be generalized to other online learning environments, such as massive open online courses. Second, we highlight the challenge of measuring SRL using trace data in online learning environments. SRL is a multifaceted concept encompassing a variety of cognitive, metacognitive, emotional, and motivational aspects, which are sometimes internal to students and difficult to measure directly (Greene & Azevedo, 2010; Winne & Perry, 2000). Consequently, no single measurement method or construct can capture all dimensions of SRL, necessitating the use of more diverse methods for its measurement and conceptualization. Our study focuses on investigating students' engagement in learning patterns within an SRL context in terms of the sequences of actions they engage in during learning. Although our method of investigating SRL-relevant strategies is not capable of capturing all aspects related to SRL, we argue that students' choices regarding the order of engaging in learning activities, as demonstrated by sequences of actions, can still provide insights into their potential use of SRL-relevant strategies.

Lastly, we acknowledge that there are limitations when potentially associating each learning sequence with corresponding SRL-relevant strategies. Each learning sequence (e.g., *Quiz* → *Quiz*) identified using sequential pattern mining might imply more than one possible SRL-relevant strategy than we associated in the study. For instance, a frequent learning pattern, *Quiz* → *Quiz*, which we categorize as a potential SRL-relevant strategy of rehearsing and memorizing, might also imply other SRL-relevant strategies, such as those related to self-evaluation. Altogether, despite the limitations we discussed, we argue that our study contributes to the discovery of the heterogeneous nature of learning patterns within an SRL context in computer-based learning environments.

Conclusion

SRL skills are invaluable in online educational environments, yet much remains to be discovered regarding what drives differences in what SRL skills are most relevant and helpful in which context. This paper contributes to enriching our understanding of the heterogeneous nature of learning patterns in an SRL context via complementary data-driven and theory-informed methods, revealing areas where SRL theories could be enhanced by data-driven insights, while also demonstrating the potential for integrating theoretical insights into data-driven AI systems for teaching SRL skills via a prototype SRL training intervention. Lastly, our findings suggest that understanding learning patterns across different subtopics can inform the design of interventions tailored to the specific characteristics of each subtopic, thereby enhancing learning outcomes.

References

Aleven, V., & Koedinger, K. R. (2002). An effective metacognitive strategy: Learning by doing and explaining with a computer-based Cognitive Tutor. *Cognitive Science*, *26*(2), 147–179. https://doi.org/10.1207/s15516709cog2602_1

Alexander, P. A., Dinsmore, D. L., Parkinson, M. M., & Winters, F. I. (2011). Self-regulated learning in academic domains. In *Handbook of Self-Regulation of Learning and Performance* (1st ed., pp. 393–407). Routledge/Taylor & Francis Group.

Alphen, E. van, & Bakker, S. (2016). Lernanto: Using an ambient display during differentiated instruction. *Proceedings of the 2016 CHI Conference Extended Abstracts on Human Factors in Computing Systems*, 2334–2340. https://doi.org/10.1145/2851581.2892524

Alvarez, R. P., Jivet, I., Perez-Sanagustin, M., Scheffel, M., & Verbert, K. (2022). Tools designed to support self-regulated learning in online learning environments: A systematic review. *IEEE Transactions on Learning Technologies*, *15*(4), 508–522. https://doi.org/10.1109/TLT.2022.3193271

Araka, E., Maina, E., Gitonga, R., & Oboko, R. (2020). Research trends in measurement and intervention tools for self-regulated learning for e-learning environments—Systematic review (2008–2018). *Research and Practice in Technology Enhanced Learning*, *15*(6). https://doi.org/10.1186/s41039-020-00129-5

Artino, A. R., & Stephens, J. M. (2009). Academic motivation and self-regulation: A comparative analysis of undergraduate and graduate students learning online. *The Internet and Higher Education*, *12*(3–4), 146–151. https://doi.org/10.1016/j.iheduc.2009.02.001

Azevedo, R. (2005). Using hypermedia as a metacognitive tool for enhancing student learning? The role of self-regulated learning. *Educational Psychologist*, *40*(4), 199–209. https://doi.org/10.1207/s15326985ep4004_2

Azevedo, R. (2014). Issues in dealing with sequential and temporal characteristics of self- and socially-regulated learning. *Metacognition and Learning*, *9*(2), 217–228. https://doi.org/10.1007/s11409-014-9123-1

Azevedo, R. (2015). Defining and measuring engagement and learning in science: Conceptual, theoretical, methodological, and analytical issues. *Educational Psychologist*, *50*(1), 84–94. https://doi.org/10.1080/00461520.2015.1004069

Azevedo, R., Moos, D. C., Johnson, A. M., & Chauncey, A. D. (2010). Measuring cognitive and metacognitive regulatory processes during hypermedia learning: Issues and challenges. *Educational Psychologist*, *45*(4), 210–223. https://doi.org/10.1080/00461520.2010.515934

Bannert, M., Reimann, P., & Sonnenberg, C. (2014). Process mining techniques for analysing patterns and strategies in students' self-regulated learning. *Metacognition and Learning*, *9*(2), 161–185. https://doi.org/10.1007/s11409-013-9107-6

Barnard, L., Paton, V., & Lan, W. (2008). Online self-regulatory learning behaviors as a mediator in the relationship between online course perceptions with achievement. *The International Review of Research in Open and Distributed Learning*, *9*(2). https://doi.org/10.19173/irrodl.v9i2.516

Barnard, L., Paton, V. O., & Lan, W. (2010). Profiles in self-regulated learning in the online learning environment. *International Review of Research in Open and Distributed Learning*, *11*(1), 61–80. https://doi.org/10.19173/irrodl.v11i1.769

Bellhäuser, H., Dignath, C., & Theobald, M. (2023). Daily automated feedback enhances self-regulated learning: A longitudinal randomized field experiment. *Frontiers in Psychology*, *14*(1125873). https://doi.org/10.3389/fpsyg.2023.1125873

Bernacki, M. L., Nokes-Malach, T. J., & Aleven, V. (2015). Examining self-efficacy during learning: Variability and relations to behavior, performance, and learning. *Metacognition and Learning*, *10*(1), 99–117. https://doi.org/10.1007/s11409-014-9127-x

Bernacki, M. L., Vosicka, L., & Utz, J. C. (2020). Can a brief, digital skill training intervention help undergraduates "learn to learn" and improve their STEM achievement? *Journal of Educational Psychology*, *112*(4), 765–781. https://doi.org/10.1037/edu0000405

Bodily, R., Ikahihifo, T. K., Mackley, B., & Graham, C. R. (2018). The design, development, and implementation of student-facing learning analytics dashboards. *Journal of Computing in Higher Education*, *30*(3), 572–598. https://doi.org/10.1007/s12528-018-9186-0

Bogarín, A., Cerezo, R., & Romero, C. (2018). Discovering learning processes using Inductive Miner: A case study with Learning Management Systems (LMSs). *Psicothema*, *30*(3), 322–329. https://doi.org/10.7334/psicothema2018.116

Bol, L., & Garner, J. K. (2011). Challenges in supporting self-regulation in distance education environments. *Journal of Computing in Higher Education*, *23*(2), 104–123. https://doi.org/10.1007/s12528-011-9046-7

Bosch, N., Zhang, Y., Paquette, L., Baker, R. S., Ocumpaugh, J., & Biswas, G. (2021). Students' verbalized metacognition during computerized learning. *Proceedings of the 2021 CHI Conference on Human Factors in Computing Systems*, *680*, 1–12. https://doi.org/10.1145/3411764.3445809

Broadbent, J., Panadero, E., Lodge, J. M., & de Barba, P. (2020). Technologies to enhance self-regulated learning in online and computer-mediated learning environments. In M. J. Bishop, E. Boling, J. Elen, & V. Svihla (Eds.), *Handbook of Research in Educational Communications and Technology: Learning Design* (pp. 37–52). Springer International Publishing. https://doi.org/10.1007/978-3-030-36119-8_3

Broadbent, J., & Poon, W. L. (2015). Self-regulated learning strategies & academic achievement in online higher education learning environments: A systematic review. *The Internet and Higher Education*, *27*, 1–13. https://doi.org/10.1016/j.iheduc.2015.04.007

Brusilovsky, P. (2001). Adaptive hypermedia. *User Modeling and User-Adapted Interaction*, *11*, 87–110. https://doi.org/10.1023/A:1011143116306

Bull, S., Johnson, M. D., Epp, C. D., Masci, D., Alotaibi, M., & Girard, S. (2014). Formative assessment and meaningful learning analytics. *Proceedings of the 2014 IEEE 14th International Conference on Advanced Learning Technologies*, 327–329. https://doi.org/10.1109/ICALT.2014.100

Carroll, C., & Garavalia, L. (2002). Gender and racial differences in select determinants of student success. *American Journal of Pharmaceutical Education*, *66*, 382–387.

Choi, H., Winne, P. H., Brooks, C., Li, W., & Shedden, K. (2023). Logs or self-reports? Misalignment between behavioral trace data and surveys when modeling learner achievement goal orientation. *Proceedings of the 13th International Conference on Learning Analytics & Knowledge*, 11–21. https://doi.org/10.1145/3576050.3576052

Cicchinelli, A., Veas, E., Pardo, A., Pammer-Schindler, V., Fessl, A., Barreiros, C., & Lindstädt, S. (2018). Finding traces of self-regulated learning in activity streams. *Proceedings of the*

8th International Conference on Learning Analytics & Knowledge, 191–200. https://doi.org/10.1145/3170358.3170381

Cloude, E. B., Azevedo, R., Winne, P. H., Biswas, G., & Jang, E. E. (2022). System design for using multimodal trace data in modeling self-regulated learning. *Frontiers in Education*, *7*(928632). https://doi.org/10.3389/feduc.2022.928632

Corrin, L., De Barba, P. G., & Bakharia, A. (2017). Using learning analytics to explore help-seeking learner profiles in MOOCs. *Proceedings of the 7th International Learning Analytics & Knowledge*, 424–428. https://doi.org/10.1145/3027385.3027448

Du, J., Hew, K. F., & Liu, L. (2023). What can online traces tell us about students' self-regulated learning? A systematic review of online trace data analysis. *Computers & Education*, *201*(104828). https://doi.org/10.1016/j.compedu.2023.104828

Du, J., & Hew, K. F. T. (2022). Using recommender systems to promote self-regulated learning in online education settings: Current knowledge gaps and suggestions for future research. *Journal of Research on Technology in Education*, *54*(4), 557–580. https://doi.org/10.1080/15391523.2021.1897905

Edisherashvili, N., Saks, K., Pedaste, M., & Leijen, Ä. (2022). Supporting self-regulated learning in distance learning contexts at higher education level: Systematic literature review. *Frontiers in Psychology*, *12*(792422). https://doi.org/10.3389/fpsyg.2021.792422

Efklides, A. (2011). Interactions of metacognition with motivation and affect in self-regulated learning: The MASRL model. *Educational Psychologist*, *46*(1), 6–25. https://doi.org/10.1080/00461520.2011.538645

Engelmann, K., Bannert, M., & Melzner, N. (2021). Do self-created metacognitive prompts promote short- and long-term effects in computer-based learning environments? *Research* *and Practice in Technology Enhanced Learning*, *16*(3). https://doi.org/10.1186/s41039-021-00148-w

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., Matcha, W., Uzir, N. A., Wang, Q., & Gašević, D. (2021). Learning analytics to reveal links between learning design and self-regulated learning. *International Journal of Artificial Intelligence in Education*, *31*(4), 980–1021. https://doi.org/10.1007/s40593-021-00249-z

Ferreira da Rocha, F. D., Lemos, B., Henrique de Brito, P., Santos, R., Rodrigues, L., Isotani, S., & Dermeval, D. (2023). Gamification and open learner model: An experimental study on the effects on self-regulatory learning characteristics. *Education and Information Technologies*, *29*, 3525–3546. https://doi.org/10.1007/s10639-023-11906-2

Foong, C. C., Bashir Ghouse, N. L., Lye, A. J., Khairul Anhar Holder, N. A., Pallath, V., Hong, W.-H., Sim, J. H., & Vadivelu, J. (2021). A qualitative study on self-regulated learning among high performing medical students. *BMC Medical Education*, *21*(320). https://doi.org/10.1186/s12909-021-02712-w

Gašević, D., Dawson, S., & Siemens, G. (2015). Let's not forget: Learning analytics are about learning. *TechTrends*, *59*(1), 64–71. https://doi.org/10.1007/s11528-014-0822-x

Graesser, A., & McNamara, D. (2010). Self-regulated learning in learning environments with pedagogical agents that interact in natural language. *Educational Psychologist*, *45*(4), 234–244. https://doi.org/10.1080/00461520.2010.515933

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., Bolick, C. M., Caprino, A. M., Deekens, V. M., McVea, M., Yu, S., & Jackson, W. P. (2015). Fostering high-school students' self-regulated learning online and across academic domains. *The High School Journal*, *99*(1), 88–106.

Greene, J. A., Bolick, C. M., Jackson, W. P., Caprino, A. M., Oswald, C., & McVea, M. (2015). Domain-specificity of self-regulated learning processing in science and history. *Contemporary Educational Psychology*, *42*, 111–128. https://doi.org/10.1016/j.cedpsych.2015.06.001

Greene, J. A., Bolick, C. M., & Robertson, J. (2010). Fostering historical knowledge and thinking skills using hypermedia learning environments: The role of self-regulated learning. *Computers & Education*, *54*(1), 230–243. https://doi.org/10.1016/j.compedu.2009.08.006

Guerra, J., Hosseini, R., Somyurek, S., & Brusilovsky, P. (2016). An intelligent interface for learning content: Combining an open learner model and social comparison to support self-regulated learning and engagement. *Proceedings of the 21st International Conference on Intelligent User Interfaces*, 152–163. https://doi.org/10.1145/2856767.2856784

Guerra, J., Schunn, C. D., Bull, S., Barria-Pineda, J., & Brusilovsky, P. (2018). Navigation support in complex open learner models: Assessing visual design alternatives. *New Review of Hypermedia and Multimedia*, *24*(3), 160–192. https://doi.org/10.1080/13614568.2018.1482375

Hadwin, A. F., Nesbit, J. C., Jamieson-Noel, D., Code, J., & Winne, P. H. (2007). Examining trace data to explore self-regulated learning. *Metacognition and Learning*, *2*, 107–124. https://doi.org/10.1007/s11409-007-9016-7

Han, F. (2023). Level of consistency between students' self-reported and observed study approaches in flipped classroom courses: How does it influence students' academic learning outcomes? *PLOS ONE*, *18*(6), e0286549. https://doi.org/10.1371/journal.pone.0286549

Han, J., Pei, J., & Tong, H. (2022). *Data mining: Concepts and techniques* (4th ed.). Morgan Kaufmann.

Hardy, J. H., Day, E. A., & Steele, L. M. (2019). Interrelationships among self-regulated learning processes: Toward a dynamic process-based model of self-regulated learning. *Journal of Management*, *45*(8), 3146–3177. https://doi.org/10.1177/0149206318780440

Heikkinen, S., Saqr, M., Malmberg, J., & Tedre, M. (2023). Supporting self-regulated learning with learning analytics interventions – a systematic literature review. *Education and Information Technologies*, *28*(3), 3059–3088. https://doi.org/10.1007/s10639-022-11281-4

Hilpert, J. C., Greene, J. A., & Bernacki, M. (2023). Leveraging complexity frameworks to refine theories of engagement: Advancing self-regulated learning in the age of artificial intelligence. *British Journal of Educational Technology*, *54*(5), 1204–1221. https://doi.org/10.1111/bjet.13340

Hooshyar, D., Pedaste, M., Saks, K., Leijen, Ä., Bardone, E., & Wang, M. (2020). Open learner models in supporting self-regulated learning in higher education: A systematic literature review. *Computers & Education*, *154*(103878). https://doi.org/10.1016/j.compedu.2020.103878

Hsiao, I.-H., Pandhalkudi Govindarajan, S. K., & Lin, Y.-L. (2016). Semantic visual analytics for today's programming courses. *Proceedings of the 6th International Conference on Learning Analytics & Knowledge*, 48–53. https://doi.org/10.1145/2883851.2883915

Irfan, M., Kusumaningrum, B., Yulia, Y., & Widodo, S. A. (2020). Challenges during the pandemic: Use of e-learning in mathematics learning in higher education. *Infinity*, *9*(2), 147–158. https://doi.org/10.22460/infinity.v9i2.p147-158

Jansen, R. S., van Leeuwen, A., Janssen, J., Conijn, R., & Kester, L. (2020). Supporting learners' self-regulated learning in Massive Open Online Courses. *Computers & Education*, *146*(103771). https://doi.org/10.1016/j.compedu.2019.103771

Järvelä, S., Järvenoja, H., & Malmberg, J. (2019). Capturing the dynamic and cyclical nature of regulation: Methodological Progress in understanding socially shared regulation in learning. *International Journal of Computer-Supported Collaborative Learning*, *14*(4), 425–441. https://doi.org/10.1007/s11412-019-09313-2

Jin, S.-H., Im, K., Yoo, M., Roll, I., & Seo, K. (2023). Supporting students' self-regulated learning in online learning using artificial intelligence applications. *International Journal of Educational Technology in Higher Education*, *20*(37). https://doi.org/10.1186/s41239-023-00406-5

Johnson, A. M., Azevedo, R., & D'Mello, S. K. (2011). The temporal and dynamic nature of self-regulatory processes during independent and externally assisted hypermedia learning. *Cognition and Instruction*, *29*(4), 471–504. https://doi.org/10.1080/07370008.2011.610244

Jovanović, J., Gašević, D., Dawson, S., Pardo, A., & Mirriahi, N. (2017). Learning analytics to unveil learning strategies in a flipped classroom. *The Internet and Higher Education*, *33*, 74–85. https://doi.org/10.1016/j.iheduc.2017.02.001

Kang, J., & Liu, M. (2022). Investigating navigational behavior patterns of students across at-risk categories within an open-ended serious game. *Technology, Knowledge and Learning*, *27*, 183–205. https://doi.org/10.1007/s10758-020-09462-6

Kang, J., Liu, M., & Qu, W. (2017). Using gameplay data to examine learning behavior patterns in a serious game. *Computers in Human Behavior*, *72*, 757–770. https://doi.org/10.1016/j.chb.2016.09.062

Kay, J., Bartimote, K., Kitto, K., Kummerfeld, B., Liu, D., & Reimann, P. (2022). Enhancing learning by Open Learner Model (OLM) driven data design. *Computers and Education: Artificial Intelligence*, *3*(100069). https://doi.org/10.1016/j.caeai.2022.100069

Kinnebrew, J. S., & Biswas, G. (2012). Identifying learning behaviors by contextualizing differential sequence mining with action features and performance evolution. *Proceedings of the 5th International Conference on Educational Data Mining*, 57–64.

Kizilcec, R. F., Pérez-Sanagustín, M., & Maldonado, J. J. (2017). Self-regulated learning strategies predict learner behavior and goal attainment in Massive Open Online Courses. *Computers & Education*, *104*, 18–33. https://doi.org/10.1016/j.compedu.2016.10.001

Kizilcec, R. F., & Schneider, E. (2015). Motivation as a lens to understand online learners: Toward data-driven design with the OLEI scale. *ACM Transactions on Computer-Human Interaction*, *22*(2), 1–24. https://doi.org/10.1145/2699735

Klug, J., Ogrin, S., Keller, S., Ihringer, A., & Schmitz, B. (2011). A plea for self-regulated learning as a process: Modelling, measuring and intervening. *Psychological Test and Assessment Modeling*, *53*(1), 51–72.

Kramarski, B., & Gutman, M. (2006). How can self-regulated learning be supported in mathematical E-learning environments?: Self-regulated learning in mathematical E-learning. *Journal of Computer Assisted Learning*, *22*, 24–33. https://doi.org/10.1111/j.1365-2729.2006.00157.x

Kuvalja, M., Verma, M., & Whitebread, D. (2014). Patterns of co-occurring non-verbal behaviour and self-directed speech; a comparison of three methodological approaches. *Metacognition and Learning*, *9*(2), 87–111. https://doi.org/10.1007/s11409-013-9106-7

Lajoie, S. P., & Azevedo, R. (2006). Teaching and learning in technology-rich environments. In *Handbook of Educational Psychology* (2nd ed., pp. 803–821). Lawrence Erlbaum Associates Publishers.

Law, C.-Y., Grundy, J., Cain, A., Vasa, R., & Cummaudo, A. (2017). User perceptions of using an open learner model visualisation tool for facilitating self-regulated learning. *Proceedings of the 19th Australasian Computing Education Conference*, 55–64. https://doi.org/10.1145/3013499.3013502

Lee, M., Lee, S. Y., Kim, J. E., & Lee, H. J. (2023). Domain-specific self-regulated learning interventions for elementary school students. *Learning and Instruction*, *88*(101810). https://doi.org/10.1016/j.learninstruc.2023.101810

Li, Q., Baker, R. B., & Warschauer, M. (2020). Using clickstream data to measure, understand, and support self-regulated learning in online courses. *The Internet and Higher Education*, *45*(100727). https://doi.org/10.1016/j.iheduc.2020.100727

Li, S., Du, H., Xing, W., Zheng, J., Chen, G., & Xie, C. (2020). Examining temporal dynamics of self-regulated learning behaviors in STEM learning: A network approach. *Computers & Education*, *158*(103987). https://doi.org/10.1016/j.compedu.2020.103987

Li, T., Fan, Y., Tan, Y., Wang, Y., Singh, S., Li, X., Raković, M., van der Graaf, J., Lim, L., Yang, B., Molenaar, I., Bannert, M., Moore, J., Swiecki, Z., Tsai, Y.-S., Shaffer, D. W., & Gašević, D. (2023). Analytics of self-regulated learning scaffolding: Effects on learning processes. *Frontiers in Psychology*, *14*(1206696). https://doi.org/10.3389/fpsyg.2023.1206696

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

Mason, L., Boldrin, A., & Ariasi, N. (2010). Epistemic metacognition in context: Evaluating and learning online information. *Metacognition and Learning*, *5*, 67–90. https://doi.org/10.1007/s11409-009-9048-2

Matcha, W., Uzir, N. A., Gašević, D., & Pardo, A. (2020). A systematic review of empirical studies on learning analytics dashboards: A self-regulated learning perspective. *IEEE Transactions on Learning Technologies*, *13*(2), 226–245. IEEE Transactions on Learning Technologies. https://doi.org/10.1109/TLT.2019.2916802

Mejia, C., Florian, B., Vatrapu, R., Bull, S., Gomez, S., & Fabregat, R. (2017). A novel web-based approach for visualization and inspection of reading difficulties on university students. *IEEE Transactions on Learning Technologies*, *10*(1), 53–67. https://doi.org/10.1109/TLT.2016.2626292

Mirriahi, N., Liaqat, D., Dawson, S., & Gašević, D. (2016). Uncovering student learning profiles with a video annotation tool: Reflective learning with and without instructional norms. *Educational Technology Research and Development*, *64*, 1083–1106. https://doi.org/10.1007/s11423-016-9449-2

Molenaar, I., Mooij, S. de, Azevedo, R., Bannert, M., Järvelä, S., & Gašević, D. (2023). Measuring self-regulated learning and the role of AI: Five years of research using multimodal multichannel data. *Computers in Human Behavior*, *139*(107540). https://doi.org/10.1016/j.chb.2022.107540

Muldner, K., Wixon, M., Rai, D., Burleson, W., Woolf, B., & Arroyo, I. (2015). Exploring the impact of a learning dashboard on student affect. *Proceedings of the 17th International Conference on Artificial Intelligence in Education*, 307–317. https://doi.org/10.1007/978-3-319-19773-9_31

Müller, N. M., & Seufert, T. (2018). Effects of self-regulation prompts in hypermedia learning on learning performance and self-efficacy. *Learning and Instruction*, *58*, 1–11. https://doi.org/10.1016/j.learninstruc.2018.04.011

Munshi, A., Rajendran, R., Ocumpaugh, J., Biswas, G., Baker, R. S., & Paquette, L. (2018). Modeling learners' cognitive and affective states to scaffold SRL in open-ended learning environments. *Proceedings of the 26th Conference on User Modeling, Adaptation and Personalization*, 131–138. https://doi.org/10.1145/3209219.3209241

Muwonge, C. M., Ssenyonga, J., Kibedi, H., & Schiefele, U. (2020). Use of self-regulated learning strategies among teacher education students: A latent profile analysis. *Social Sciences & Humanities Open*, *2*(1), 100037. https://doi.org/10.1016/j.ssaho.2020.100037

Ng, J. T. D., Liu, Y., Chui, D. S. Y., Man, J. C. H., & Hu, X. (2023). Leveraging LMS logs to analyze self-regulated learning behaviors in a maker-based course. *Proceedings of the 13th International Conference on Learning Analytics and Knowledge Conference*, 670–676. https://doi.org/10.1145/3576050.3576111

Norman, E., & Furnes, B. (2016). The relationship between metacognitive experiences and learning: Is there a difference between digital and non-digital study media? *Computers in Human Behavior*, *54*, 301–309. https://doi.org/10.1016/j.chb.2015.07.043

Palanci, A., Yılmaz, R. M., & Turan, Z. (2024). Learning analytics in distance education: A systematic review study. *Education and Information Technologies*. https://doi.org/10.1007/s10639-024-12737-5

Panadero, E. (2017). A review of self-regulated learning: Six models and four directions for research. *Frontiers in Psychology*, *8*(422). https://doi.org/10.3389/fpsyg.2017.00422

Panadero, E., Klug, J., & Järvelä, S. (2016). Third wave of measurement in the self-regulated learning field: When measurement and intervention come hand in hand. *Scandinavian Journal of Educational Research*, *60*(6), 723–735. https://doi.org/10.1080/00313831.2015.1066436

Paquette, L., Grant, T., Zhang, Y., Biswas, G., & Baker, R. S. (2021). Using epistemic networks to analyze self-regulated learning in an open-ended problem-solving environment. *Proceedings of the 2nd International Conference on Quantitative Enthnography*, 185–201. https://doi.org/10.1007/978-3-030-67788-6_13

Pedrotti, M., & Nistor, N. (2019). How students fail to self-regulate their online learning experience. *Proceedings of the 14th European Conference on Technology Enhanced Learning*, 377–385. https://doi.org/10.1007/978-3-030-29736-7_28

Peer, E., Rothschild, D., Gordon, A., Evernden, Z., & Damer, E. (2021). Data quality of platforms and panels for online behavioral research. *Behavior Research Methods*, *54*(4), 1643–1662. https://doi.org/10.3758/s13428-021-01694-3

Perera, D., Kay, J., Koprinska, I., Yacef, K., & Zaïane, O. R. (2009). Clustering and sequential pattern mining of online collaborative learning data. *IEEE Transactions on Knowledge and Data Engineering*, *21*(6), 759–772. https://doi.org/10.1109/TKDE.2008.138

Pieger, E., & Bannert, M. (2018). Differential effects of students' self-directed metacognitive prompts. *Computers in Human Behavior*, *86*, 165–173. https://doi.org/10.1016/j.chb.2018.04.022

Pintrich, P. R. (2000). The role of goal orientation in self-regulated learning. In *Handbook of Self-Regulation* (pp. 451–502). Academic Press. https://doi.org/10.1016/B978-012109890-2/50043-3

Pintrich, P. R., Smith, D., García, T., & McKeachie, W. (1991). *A manual for the use of the Motivated Strategies for Learning Questionnaire (MSLQ)* (NCRIPTAL-91-B-004). National Center for Research to Improve Postsecondary Teaching and Learning, Ann Arbor, MI. https://eric.ed.gov/?id=ED338122

Pinxteren, S., & Calders, T. (2021). Efficient permutation testing for significant sequential patterns. *Proceedings of the 2021 SIAM International Conference on Data Mining*, 19–27. https://doi.org/10.1137/1.9781611976700.3

Poitras, E. G., & Lajoie, S. P. (2013). A domain-specific account of self-regulated learning: The cognitive and metacognitive activities involved in learning through historical inquiry. *Metacognition and Learning*, *8*(3), 213–234. https://doi.org/10.1007/s11409-013-9104-9

Poon, L. K. M., Kong, S.-C., Wong, M. Y. W., & Yau, T. S. H. (2017). Mining sequential patterns of students' access on learning management system. *Proceedings of the 2nd International Conference on Data Mining and Big Data*, 191–198. https://doi.org/10.1007/978-3-319-61845-6_20

Puustinen, M., & Pulkkinen, L. (2001). Models of self-regulated learning: A review. *Scandinavian Journal of Educational Research*, *45*(3), 269–286. https://doi.org/10.1080/00313830120074206

Reimann, P. (2009). Time is precious: Variable- and event-centred approaches to process analysis in CSCL research. *International Journal of Computer-Supported Collaborative Learning*, *4*(3), 239–257. https://doi.org/10.1007/s11412-009-9070-z

Richardson, M., Abraham, C., & Bond, R. (2012). Psychological correlates of university students' academic performance: A systematic review and meta-analysis. *Psychological Bulletin*, *138*(2), 353–387. https://doi.org/10.1037/a0026838

Rowe, E., Asbell-Clarke, J., & Baker, R. S. (2015). Serious games analytics to measure implicit science learning. In C. S. Loh, Y. Sheng, & D. Ifenthaler (Eds.), *Serious Games Analytics* (pp. 343–360). Springer International Publishing. https://doi.org/10.1007/978-3-319-05834-4_15

Saint, J., Fan, Y., Singh, S., Gasevic, D., & Pardo, A. (2021). Using process mining to analyse self-regulated learning: A systematic analysis of four algorithms. *Proceedings of the 11th International Conference on Learning Analytics & Knowledge*, 333–343. https://doi.org/10.1145/3448139.3448171

Schraw, G., Crippen, K. J., & Hartley, K. (2006). Promoting self-regulation in science education: Metacognition as part of a broader perspective on learning. *Research in Science Education*, *36*, 111–139. https://doi.org/10.1007/s11165-005-3917-8

Schunk, D. H., & Greene, J. A. (2017). *Handbook of self-regulation of learning and performance* (2nd ed.). Routledge.

Schunk, D. H., & Zimmerman, B. J. (1997). Social origins of self-regulatory competence. *Educational Psychologist*, *32*(4), 195–208. https://doi.org/10.1207/s15326985ep3204_1

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

Shaffer, D. W. (2004). Epistemic frames and islands of expertise: Learning from infusion experiences. *Proceedings of the 6th International Conference on Learning Sciences*, 473–480.

Shaffer, D. W. (2006). Epistemic frames for epistemic games. *Computers & Education*, *46*(3), 223–234. https://doi.org/10.1016/j.compedu.2005.11.003

Shirvani Boroujeni, M., & Dillenbourg, P. (2019). Discovery and temporal analysis of MOOC study patterns. *Journal of Learning Analytics*, *6*(1). https://doi.org/10.18608/jla.2019.61.2

Siadaty, M., Gašević, D., & Hatala, M. (2016). Measuring the impact of technological scaffolding interventions on micro-level processes of self-regulated workplace learning. *Computers in Human Behavior*, *59*, 469–482. https://doi.org/10.1016/j.chb.2016.02.025

Siemens, G., & Baker, R. S. (2012). Learning analytics and educational data mining: Towards communication and collaboration. *Proceedings of the 2nd International Conference on Learning Analytics and Knowledge*, 252–254. https://doi.org/10.1145/2330601.2330661

Sobocinski, M., Malmberg, J., & Järvelä, S. (2017). Exploring temporal sequences of regulatory phases and associated interactions in low- and high-challenge collaborative learning sessions. *Metacognition and Learning*, *12*(2), 275–294. https://doi.org/10.1007/s11409-016-9167-5

Sonnenberg, C., & Bannert, M. (2015). Discovering the effects of metacognitive prompts on the sequential structure of SRL-processes using process mining techniques. *Journal of Learning Analytics*, *2*(1). https://doi.org/10.18608/jla.2015.21.5

Sonnenberg, C., & Bannert, M. (2019). Using Process Mining to examine the sustainability of instructional support: How stable are the effects of metacognitive prompting on self-regulatory behavior? *Computers in Human Behavior*, *96*, 259–272. https://doi.org/10.1016/j.chb.2018.06.003

Sun, J. C.-Y., Tsai, H.-E., & Cheng, W. K. R. (2023). Effects of integrating an open learner model with AI-enabled visualization on students' self-regulation strategies usage and behavioral patterns in an online research ethics course. *Computers and Education: Artificial Intelligence*, *4*(100120). https://doi.org/10.1016/j.caeai.2022.100120

Syal, S., & Nietfeld, J. L. (2020). The impact of trace data and motivational self-reports in a game-based learning environment. *Computers & Education*, *157*(103978). https://doi.org/10.1016/j.compedu.2020.103978

Tacoma, S., Sosnovsky, S., Boon, P., Jeuring, J., & Drijvers, P. (2018). The interplay between inspectable student models and didactics of statistics. *Digital Experiences in Mathematics Education*, *4*(2), 139–162. https://doi.org/10.1007/s40751-018-0040-9

Taub, M., Azevedo, R., Rajendran, R., Cloude, E. B., Biswas, G., & Price, M. J. (2021). How are students' emotions related to the accuracy of cognitive and metacognitive processes during learning with an intelligent tutoring system? *Learning and Instruction*, *72*(101200). https://doi.org/10.1016/j.learninstruc.2019.04.001

Tonon, A., & Vandin, F. (2019). Permutation strategies for mining significant sequential patterns. *Proceedings of the 2019 IEEE 19th International Conference on Data Mining*, 1330–1335. https://doi.org/10.1109/ICDM.2019.00169

Tseng, W.-T., Dörnyei, Z., & Schmitt, N. (2006). A new approach to assessing strategic learning: The case of self-regulation in vocabulary acquisition. *Applied Linguistics*, *27*(1), 78–102. https://doi.org/10.1093/applin/ami046

van Alten, D. C. D., Phielix, C., Janssen, J., & Kester, L. (2020). Self-regulated learning support in flipped learning videos enhances learning outcomes. *Computers & Education*, *158*(104000). https://doi.org/10.1016/j.compedu.2020.104000

van der Graaf, J., Lim, L., Fan, Y., Kilgour, J., Moore, J., Gašević, D., Bannert, M., & Molenaar, I. (2022). The dynamics between self-regulated learning and learning outcomes: An exploratory approach and implications. *Metacognition and Learning*, *17*(3), 745–771. https://doi.org/10.1007/s11409-022-09308-9

Veenman, M. V. J. (2016). Learning to self-monitor and self-regulate. In *Handbook of Research on Learning and Instruction* (2nd ed.). Routledge.

Veenman, M. V. J., & van Cleef, D. (2019). Measuring metacognitive skills for mathematics: Students' self-reports versus on-line assessment methods. *ZDM*, *51*(4), 691–701. https://doi.org/10.1007/s11858-018-1006-5

Virtanen, P., & Nevgi, A. (2010). Disciplinary and gender differences among higher education students in self-regulated learning strategies. *Educational Psychology*, *30*(3), 323–347. https://doi.org/10.1080/01443411003606391

Winne, P. H. (2010). Improving measurements of self-regulated learning. *Educational Psychologist*, *45*(4), 267–276. https://doi.org/10.1080/00461520.2010.517150

Winne, P. H. (2021). Open Learner Models working in symbiosis with self-regulating learners: A research agenda. *International Journal of Artificial Intelligence in Education*, *31*(3), 446–459. https://doi.org/10.1007/s40593-020-00212-4

Winne, P. H., & Hadwin, A. F. (1998). Studying as self-regulated learning. In D. J. Hacker & J. Dunlosky (Eds.), *Metacognition in Educational Theory and Practice* (1st ed., pp. 277–304). Lawrence Erlbaum.

Winne, P. H., & Jamieson-Noel, D. (2002). Exploring students' calibration of self reports about study tactics and achievement. *Contemporary Educational Psychology*, *27*(4), 551–572. https://doi.org/10.1016/S0361-476X(02)00006-1

Winne, P. H., & Perry, N. E. (2000). Measuring self-regulated learning. In *Handbook of Self-Regulation* (pp. 531–566). Elsevier. https://doi.org/10.1016/B978-012109890-2/50045-7

Winters, F. I., Greene, J. A., & Costich, C. M. (2008). Self-regulation of learning within computer-based learning environments: A critical analysis. *Educational Psychology Review*, *20*(4), 429–444. https://doi.org/10.1007/s10648-008-9080-9

Wong, J., Khalil, M., Baars, M., de Koning, B. B., & Paas, F. (2019). Exploring sequences of learner activities in relation to self-regulated learning in a massive open online course. *Computers & Education*, *140*(103595). https://doi.org/10.1016/j.compedu.2019.103595

Xu, Z., Zhao, Y., Liew, J., Zhou, X., & Kogut, A. (2023). Synthesizing research evidence on self-regulated learning and academic achievement in online and blended learning environments: A scoping review. *Educational Research Review*, *39*(100510). https://doi.org/10.1016/j.edurev.2023.100510

Ye, D., & Pennisi, S. (2022). Using trace data to enhance students' self-regulation: A learning analytics perspective. *The Internet and Higher Education*, *54*(100855). https://doi.org/10.1016/j.iheduc.2022.100855

Yukselturk, E., & Bulut, S. (2009). Gender differences in self-regulated online learning environment. *Educational Technology & Society*, *12*(3), 12–22.

Zaki, M. J. (2000). Scalable algorithms for association mining. *IEEE Transactions on Knowledge and Data Engineering*, *12*(3), 372–390. https://doi.org/10.1109/69.846291

Zaki, M. J. (2001). SPADE: An efficient algorithm for mining frequent sequences. *Machine Learning*, *42*(1), 31–60. https://doi.org/10.1023/A:1007652502315

Zarei Hajiabadi, Z., Gandomkar, R., Sohrabpour, A. A., & Sandars, J. (2023). Developing low-achieving medical students' self-regulated learning using a combined learning diary and explicit training intervention. *Medical Teacher*, *45*(5), 475–484. https://doi.org/10.1080/0142159X.2022.2152664

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 Nature Singapore. https://doi.org/10.1007/978-981-99-0026-8_6

Zhang, Y., Paquette, L., Baker, R. S., Ocumpaugh, J., Bosch, N., Munshi, A., & Biswas, G. (2020). The relationship between confusion and metacognitive strategies in Betty's Brain. *Proceedings of the 10th International Conference on Learning Analytics & Knowledge*, 276–284. https://doi.org/10.1145/3375462.3375518

Zhang, Y., Paquette, L., & Bosch, N. (2024). Using permutation tests to identify statistically sound and nonredundant sequential patterns in educational event sequences. *Journal of*

Educational and Behavioral Statistics, 1–33. https://doi.org/10.3102/10769986241248772

Zhang, Y., Paquette, L., Bosch, N., Ocumpaugh, J., Biswas, G., Hutt, S., & Baker, R. S. (2022). The evolution of metacognitive strategy use in an open-ended learning environment: Do prior domain knowledge and motivation play a role? *Contemporary Educational Psychology*, *69*(102064). https://doi.org/10.1016/j.cedpsych.2022.102064

Zheng, J., Li, S., & Lajoie, S. P. (2022). Diagnosing virtual patients in a technology-rich learning environment: A sequential mining of students' efficiency and behavioral patterns. *Education and Information Technologies*, *27*(3), 4259–4275. https://doi.org/10.1007/s10639-021-10772-0

Zhichun Liu & Jewoong Moon. (2023). A framework for applying sequential data analytics to design personalized digital game-based learning for computing education. *Educational Technology & Society*, *26*(2). https://doi.org/10.30191/ETS.202304_26(2).0013

Zimmerman, B. J. (1989). A social cognitive view of self-regulated academic learning. *Journal of Educational Psychology*, *81*(3), 329–339. https://doi.org/10.1037/0022-0663.81.3.329

Zimmerman, B. J. (2002). Becoming a self-regulated learner: An overview. *Theory Into Practice*, *41*(2), 64–70. https://doi.org/10.1207/s15430421tip4102_2

Zimmerman, B. J., & Kitsantas, A. (2005). The hidden dimension of personal competence: Self-regulated learning and practice. In *Handbook of Competence and Motivation* (pp. 509–526). Guilford Publications.

Zimmerman, B. J., & Kitsantas, A. (2007). Reliability and validity of self-efficacy for learning form (SELF) scores of college students. *Zeitschrift Für Psychologie/Journal of Psychology*, *215*(3), 157–163. https://doi.org/10.1027/0044-3409.215.3.157

Zimmerman, B. J., & Moylan, A. R. (2009). Self-regulation: Where metacognition and motivation intersect. In *Handbook of Metacognition in Education* (pp. 299–315). Routledge/Taylor & Francis Group.

Zimmerman, B. J., & Pons, M. M. (1986). Development of a structured interview for assessing student use of self-regulated learning strategies. *American Educational Research Journal*, *23*(4), 614–628. https://doi.org/10.2307/1163093

Zimmerman, B. J., & Pons, M. M. (1990). Student differences in self-regulated learning: Relating grade, sex, and giftedness to self-efficacy and strategy use. *Journal of Educational Psychology*, *82*(1), 51–59. https://doi.org/10.1037/0022-0663.82.1.51

Appendix

Table 1

Distribution of Students by Race and Ethnicity

Race/ethnicity category	**Local (*n*=112)**	**Prolific (*n*=98)**	**Responses**
White	63 (56.3%)	54 (55.1%)	White, Caucasian, European/Middle East and North Africa White/Hispanic, White/Mexican
Asian	28 (25%)	16 (16.3%)	Asian, Asian American, Asian/White, Chinese American, Indian, Chinese, East Asian, Filipino, Korean American, South Asian, Middle Eastern, South Asian
Black	9 (8.04%)	12 (12.2%)	Black, African American, Black – Caribbean American, Black American
Latinx/Hispanic	9 (8.04%)	11 (11.2%)	Chicano, Hispanic, Latina(o), Mexican, Mexican American
Others	3 (2.62%)	5 (5.2%)	Prefer not to answer, Guyanese American, Pacific Islander, Native American, Black and Mexican

Table 2

Distribution of Students by Gender

Gender category	**Local (*n*=112)**	**Prolific (*n*=98)**	**Responses**
Female	83 (74.1%)	44 (44.9%)	Female, Woman, Cisgender Woman
Male	26 (23.2%)	44 (44.9%)	Male, Man, Cisgender Male

Additional (grouped for anonymity)	3 (2.7%)	10 (10.2%)	Non-binary (Non binary, they), Transgender (Transgender man, Transman, Transmasculine), Queer (Genderqueer), Prefer not to answer

Table 3

Distribution of Students by English as a First Language

English as a first language category	**Local (*n*=112)**	**Prolific (*n*=98)**
English is the only first language	84 (75.0%)	77 (78.6%)
English is among the multiple first languages	17 (15.2%)	14 (14.3%)
English is not my first language; I Prefer not to answer	11 (9.8%)	7 (7.1%)

Table 4

Distribution of Students by Age

Age category	Local (*n*=112)	Prolific (*n*=98)
18-20	84 (75%)	23 (23.5%)
21-25	24 (21.4%)	42 (42.9%)
26-30	2 (1.79%)	13 (13.4%)
31-35	1 (0.905%)	10 (10.1%)
Over 35	1 (0.905%)	10 (10.1%)

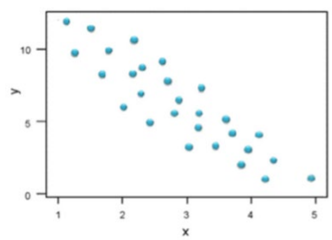
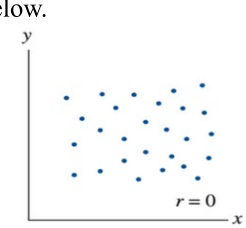
Table 5

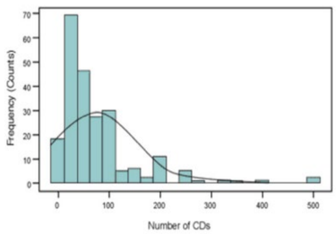
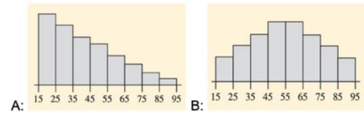
Distribution of Students by Class Standing

Class standing category	Local (*n*=112)	Prolific (*n*=98)
Freshmen	8 (7.14%)	6 (6.12%)
Sophomore	54 (48.2%)	15 (15.3%)
Junior	34 (9.8%)	40 (40.8%)
Senior	9 (8.04%)	30 (30.6%)
Graduate or Professional	6 (5.35%)	2 (2.06%)
Prefer not to answer	1 (0.86 %)	5 (5.12%)

Table 6

Full questions for test A and test B and the corresponding correct response rate for each question.

Test A	Test A correct response rate	Test B	Test B correct response rate
Identify each of the following variables. a. The time it takes to run a marathon b. The choice of diet (vegetarian or non-vegetarian) o a: quantitative and discrete & b: categorical and nominal o a: quantitative and discrete & b: categorical and ordinal o a: quantitative and continuous & b: categorical and nominal o a: quantitative and continuous & b: categorical and ordinal	48.8%	Identify each of the following variables. a. Number of pets in family b. Choice of auto to buy (domestic or import) o a: quantitative and discrete & b: categorical and nominal o a: quantitative and discrete & b: categorical and ordinal o a: quantitative and continuous & b: categorical and nominal o a: quantitative and continuous & b: categorical and ordinal	54.1%
Describe the association found in the graph below.  o Positive linear association o Positive nonlinear association o Negative linear association o Negative nonlinear association	65.7%	Describe the association found in the graph below.  o Positive linear association o Negative linear association o Negative nonlinear association o Neither positive nor negative linear association	60.4%
Which of the following variables will most likely to follow a normal curve? o The distribution of time to complete the course for all of the competitors in the Boston Marathon o The distribution of difference between the weight today and the weight tomorrow of the cows at a large dairy farm o The distribution of age of death from heart disease o The distribution of individual incomes in the U.S.	27.5%	Which of the following variables is most likely to follow a normal curve? o The distribution of height of female college students o The distribution of age of death from cancer o The distribution of scores on a difficult exam o The distribution of the household income in the U.S.	46.9%
The histogram below displays the number of CDs owned from a sample of STAT 100	58.5%	Describe the shapes of the following histograms.	42.0%

students. What shape is displayed on this histogram?  o Symmetric o Skewed to the right o Skewed to the left o No clear trend		 o A: skewed to the left & B: symmetric and unimodal o A: skewed to the right & B: symmetric and bimodal o A: skewed to the left & B: symmetric and bimodal o A: skewed to the right & B: symmetric and unimodal	
Choose which type of statistics describes the following. A prediction has been made that 5% of school students will participate in the oratorical contest. o Descriptive o Inferential o Survey o Percentage	88.4%	Choose which type of statistics describes the following. Previous survey found that 85% of college students do not have a car by displaying numerical summaries. o Descriptive o Inferential o Both descriptive and inferential o Percentage	48.8%
Which of the following is true about independent variables? o An independent variable is sometimes called the response variable o An independent variable is sometimes called the explanatory variable o An independent variable can be found along the y-axis o An independent variable is usually the outcome of the study	50.2%	Which of the following is true about dependent variable? o Dependent variable is sometimes called the response variable o Dependent variable is sometimes called the explanatory variable o Dependent variable can be found along the x-axis o Dependent variable is used to explain other variables	70.5%
Which data point from the following ordered values could be considered as an outlier? 0.1, 0.4, 1.5, 2.6, 13 o 0.1 o 0.4 o 1.5 o 2.6 o 13	89.4%	Given the following data set, calculate the mean of the data set without the outlier. 3, 4, 15, 2, 5, 1 o 3 o 5 o 2 o 15	86.0%
Which of the following is the range of possible values that a correlation can assume? o 0 to 1 o -1 to 0 o -1 to 1 o 0 and above	74.9%	Which of the following is true about correlation? o The value of the correlation depends on the variables' units o The correlation value can be greater than 1 o The correlation coefficient shows both the direction and the	78.7%

		strength of relationship between variables ○ Two variables do not have the same correlation no matter which is treated as the response variable and which is treated as the explanatory variable	
Generally, if mean is less than the median, the distribution is: ○ Skewed to the right ○ Skewed to the left ○ Symmetric ○ No distribution	52.7%	Generally, if mean is greater than the mode, the distribution is: ○ Skewed to the right ○ Skewed to the left ○ Symmetric ○ No distribution	41.1%
If the standard deviation of a dataset is 4, what is the variance? ○ 2 ○ 4 ○ 8 ○ 16	74.9%	If the variance of a data set is 16, what is the standard deviation? ○ 4 ○ 8 ○ 16 ○ 256	63.3%
A low standard deviation implies that data points are: ○ Clustered around the outlier ○ Distant from the outlier ○ Distant from the mean ○ Clustered around the mean	72.0%	A high standard deviation implies that data points are: ○ Clustered around the outlier ○ Distant from the outlier ○ Distant from the mean ○ Clustered around the mean	59.4%
Which of the following surveys would have the biggest margin of error? ○ A sample size of n = 10 ○ A sample size of n = 100 ○ A sample size of n = 10,000 ○ A sample size of n = 100,000	70.0%	Suppose a margin of error for a poll is 4%. What is the correct interpretation of the margin of error for this poll? In about 95% of all samples of this size, the _________. ○ Difference between the sample percent and the population percent will be within 4% ○ Probability that the sample percent does not equal the population percent is 4% ○ Probability that the sample percent does equal the population percent is 4% ○ Difference between the sample percent and the population percent will exceed 4%	68.6%

Note. Each question in test A and test B is displayed in an order that aligns with the testing of similar statistical concepts. The correct response rate refers to the percentage of students answering each question correctly. Some of the images were adapted from *Statistics: The Art and*

Science of Learning from Data, 3rd edition by Alan Agresti and Christine Franklin (Pearson Education, 2013).

Table 7

Regression Table for RQ1 Model

Dependent variable	**Independent variables**			
Sequence	Terminology	Graph	Calculation	Amalgamation
Read → Quiz	Reference variable	-.397***	-.492***	-.405***
Quiz → Read		.032	-.056	-.071
Quiz → Quiz		.120	-.167*	-.267***
Quiz → Examples		-.061	-.147	-.037
Read → Examples		-.117	.026	-.052
Quiz → Summary		-.290***	-.489***	-.360***
Read → Quiz	.400***	Reference variable	-.095	-.008
Quiz → Read	-.032		-.087	-.103
Quiz → Quiz	-.120		-.287***	-.387***
Quiz → Examples	.061		-.086	.024
Read → Examples	.117		.143	.065
Quiz → Summary	.290***		-.200*	-.069
Read → Quiz	.492***	.095	Reference variable	.087
Quiz → Read	.056	.087		-.016
Quiz → Quiz	.167*	.287***		-.100
Quiz → Examples	.147	.086		.110
Read → Examples	-.143	-.026		-.078
Quiz → Summary	.489***	.199*		.130
Read → Quiz	.405***	.008	-.087	Reference variable
Quiz → Read	.071	.103	.016	
Quiz → Quiz	.267***	.387***	.100	
Quiz → Examples	.037	-.024	-.110	
Read → Examples	.052	-.065	.078	
Quiz → Summary	.360***	.069	-.120	

Note. The term "Reference variable" implies that an independent variable was used as the reference category within the model. Therefore, in interpreting the model, the outcomes for other independent variables are evaluated in comparison to this reference variable. The levels of significance are denoted as follows: * p < .05, ** p < .01, *** p < .001.

Table 8

Regression Table for RQ2 Model

Dependent variable / **Learning pattern**	Terminology	Graph	Calculation	Amalgamation	Prior Knowledge	Learning Gain	Terminology* Learning Gain	Graph* Learning Gain	Calculation* Learning Gain	Amalgamation* Learning Gain	Terminology* Prior Knowledge	Graph* Prior Knowledge	Calculation* Prior Knowledge	Amalgamation* Prior Knowledge
Read → Quiz	Reference variable	-.408***	-.538***	-.412***	.038	.027		.046	.018	.020		.085	.071	-.078
Quiz → Read		.069	.027	-.043	-.079	-.023		-.021	-.147	-.002		.114	-.102	.090
Quiz → Quiz		.134	-.148	-.274**	-.028	.061		.118	-.234*	-.009		.276*	-.007	.092
Quiz → Examples		-.137	-.180	-.117	.150	.211**		-.160	-.289*	-.190		-.134	-.296*	-.114
Read → Examples		-.073	.058	.000	-.085	-.095		.160	.174	.057		.189	.108	.018
Quiz → Summary		-.299***	-.512***	-.358***	.015***	.053***		-.043	-.034	-.017		.007	.036	-.064
Read → Quiz	.408***	Reference variable	-.130	-.004	.123	.073	-.046		-.028	-.026	-.084		-.014	-.163
Quiz → Read	-.069		-.042	-.112	.035	-.043	.021		-.127	.018	-.114		-.215	-.023
Quiz → Quiz	-.134**		-.281***	-.408***	.248	.179*	-.118		-.352***	-.127	-.276*		-.283**	-.184
Quiz → Examples	.137		-.044	.020	.016	.051	.160		-.128	-.029	.134		-.162	.020
Read → Examples	.073		.131	.074	.104	.064	-.160		.014	-.103	-.189		-.082	-.171
Quiz → Summary	.299***		-.213*	-.059	.021	.010	.043		.009	.027	-.007		.029	-.070
Read → Quiz	.299***	.130	Reference variable	.126	.109	.044	-.018	.028		.003	-.071	.014		-.149
Quiz → Read	-.027	.042		-.070	-.181*	-.170	.147	.127		.145	.102	.215		.192
Quiz → Quiz	.148	.281***		-.126	-.035	-.173*	.234*	.352***		.225*	.007	.283**		.099
Quiz → Examples	.180	.044		.064	-.146	-.077	.289*	.128		.100	.296*	.162		.182
Read → Examples	-.058	-.131		-.057	.022	.079	-.174	-.014		-.117	-.108	.082		-.089

Independent variables: Terminology through Amalgamation* Prior Knowledge.

Quiz → Summary	.512***	.213*		.154	.051	.019	.034	-.009		.018	-.036	-.029		-.100
Read → Quiz	.413***	.004	-.126	Reference variable	-.040	.047	-.020	.026	-.003		.078	.163	.149	
Quiz → Read	.043	.112	.070		.011	-.025	.002	-.018	-.145		-.090	.023	-.192	
Quiz → Quiz	.274**	.408***	.126		.064	.052	.009	.127	-.225*		-.092	.184	-.099	
Quiz → Examples	.117	-.020	-.064		.036	.022	.189	.029	-.100		.114	-.020	-.182	
Read → Examples	.000	-.074	.057		-.067	-.038	-.057	.103	.117		-.018	.171	.089	
Quiz → Summary	.358***	.059	-.154		-.049	.037	.017	-.027	-.018		.064	.070	.100	

Note. The term “Reference variable” implies that an independent variable was used as the reference category within the model. Therefore, in interpreting the model, the outcomes for other independent variables are evaluated in comparison to this reference variable. The levels of significance are denoted as follows: * $p < .05$, ** $p < .01$, *** $p < .001$.

Table 9

Regression Table for RQ3 Model

Dependent variable	**Independent Variables**													
	Terminology	Graph	Calculation	Amalgamation	Coherent reading	Coherent quiz	Terminology* Coherent reading	Graph* Coherent reading	Calculation* Coherent reading	Amalgamation* Coherent reading	Terminology* Coherent quiz	Graph* Coherent quiz	Calculation* Coherent quiz	Amalgamation* Coherent quiz
Learning gain	Reference variable	6.971*	3.657	2.567	-.978	1.818		-.153	-.438	.718		4.854	-6.212*	-.900
	-6.971*	Reference variable	-3.314	-4.405	-1.131	6.672	.153		-.285	.872	-4.854		-11.066*	-5.755
	-3.657	3.314	Reference variable	-1.091	-1.416	-4.393*	.438	.285		1.157	6.212*	11.066*		5.311
	-2.567	4.405	1.091	Reference variable	-.259	.918	-.718	-.0872	-1.157		.900	5.755	-5.311	

Note. The term “Reference variable” implies that an independent variable was used as the reference category within the model. Therefore, in interpreting the model, the outcomes for other independent variables are evaluated in comparison to this reference variable. The levels of significance are denoted as follows: * p < .05, ** p < .01, *** p < .001.