Exploring Temporality in Learning Analytics: A Comprehensive Framework for Temporal Educational Studies

by

Sina Nazeri

MSc., (Cognitive Science and Intelligent Systems), UNIMAS, 2016

B. Eng., (Information Technology), Azad University, 2011

Thesis Submitted in Partial Fulfillment of the
Requirements for the Degree of
Doctor of Philosophy

in the
School of Interactive Arts and Technology
Faculty of Communication, Art and Technology

SIMON FRASER UNIVERSITY
Spring 2024

Declaration of Committee

Name:	**Sina Nazeri**
Degree:	**Doctor of Philosophy**
Title:	**Exploring Temporality in Learning Analytics: A Comprehensive Framework for Temporal Educational Studies**
Committee:	**Chair: Nilay Yalçin** Assistant Professor, Interactive Arts and Technology
	Marek Hatala Supervisor Professor, Interactive Arts and Technology
	Carman Neustaedter Committee Member Professor, Interactive Arts and Technology
	John Nesbit Examiner Professor, Education
	Bodong Chen External Examiner Associate Professor, Learning, Teaching, and Literacies, University of Pennsylvania

Abstract

With the increasing focus on the temporal aspects of learning as an emerging area of study in education research, there is a new need for a guiding framework to help researchers and practitioners navigate in this area. Such a framework can provide clarity, mitigate redundancy, and offer a systematic way to approach and handle the challenges of temporal analytics.

This thesis makes several contributions to the expanding field of temporal analytics, and thus, learning analytics. It encompasses two main contributions: a comprehensive mapping study and a systematically developed framework. Additionally, it offers two auxiliary contributions in the form of follow-up case studies.

The research starts with a systematic mapping study, categorizing elements of published research to understand current trends that focus on the temporal dimension in educational studies. By employing thematic coding, the study analyzed patterns in temporal research, including the theme of research questions being asked, data collection with different granular levels, analytical techniques employed, and the derived learning insights.

Building upon the mapping study, the thesis introduces a novel framework for temporal analytics. This framework aids researchers in data design, feature engineering, and analysis phases of temporal research. It also includes a set of guiding questions for each stage, enhancing the process of conducting temporal studies.

To demonstrate the framework's applicability, two case studies were conducted. The first explores the temporal nature of students’ self-regulated learning (SRL) behaviors. revealing various temporal patterns in SRL phase transitions. The second case study investigates temporality in students' posting behaviors in online discussions. It underscores the influence of visualization feedback on discussion engagement, particularly in motivating students and fostering behavioral change. Both case studies evaluate and exemplify the framework's effectiveness in providing a richer understanding of learners' temporal behaviors, going beyond traditional correlational or cross-sectional methods. The insights from these studies offer practical implications for designing effective SRL feedback systems and online discussion dashboards, illustrating the framework's adaptability across various learning contexts.

Overall, the thesis offers a comprehensive approach to temporal data analysis. Unlike existing models that focus on specific aspects, this framework offers a broader, more inclusive perspective, making it versatile for a wide range of studies in the field. The thesis also addresses the framework's limitations and potential areas for improvement.

Keywords: learning analytics; temporal analytics; framework; self-regulated learning (SRL); visualization dashboard

Acknowledgements

I would like to acknowledge and give my warmest thanks to my supervisor Dr. Marek Hatala who made this work possible. His guidance and advice carried me through all the stages of writing my project. I would also like to thank my co-supervisor member, Dr. Carman Neustaedter, for your brilliant comments and suggestions, thanks to you.

I am also grateful for the support from the Natural Sciences and Engineering Research Council of Canada (NSERC), under grant number RGPIN-2018-06071.

A special note of thanks is extended to the modern tools that assisted in refining this thesis. I am particularly grateful for Grammarly, LLAMA 2 by Meta, which was helpful in proofreading and grammar checking, ensuring the clarity and precision of the language used in this document.

To ensure integrity and avoid misinterpretation, I used these AI tools solely for proofreading purposes and tailored the following prompt accordingly:

You are a Grammar only which is a specialized model designed for proofreading and editing academic texts. Its primary role is to correct grammatical errors while making minimal changes to the content, ensuring that the original context and argument of the author are maintained. In cases where the context or meaning is unclear, you need to ask for clarification to ensure accuracy in its corrections. The output will be in a formal, academic tone, aligning with the nature of the texts it is editing.

Example input:

We start with descriptive data analysis to grasp the general trend in our data and get familiar with the data. Next we performed the lag sequential analysis (LSA) according to Bakeman’s method on the sequence analysis (Bakeman & M.Gottman, 1997).

Output:

We began with descriptive data analysis to understand the general trends in our data and familiarize ourselves with it. Next, we conducted a lag sequential analysis (LSA) following Bakeman and Gottman's approach (Bakeman & M. Gottman, 1997).

Table of Contents

List of Tables

List of Figures

Chapter 1. Introduction

1.1. Learning Analytics and Temporal Analytics

The rapid advancement and extensive adoption of technology in education media have generated copious amounts of data which can provide the knowledge needed to improve learning and education (Bienkowski et al., 2014). To fulfill this promise, the field of learning analytics (LA) is formed to expand our understanding of learning and how to improve learning (Gaševic et al., 2015). According to Zimmerman's learning definition, learning is the acquisition of knowledge that influences the thinking and the behavior of individuals (Zimmerman, 1990). As for understanding the learning phenomenon, it is very important to understand the innate relation between time and learning; learning has a temporal characteristic in nature, which means that it occurs over the passage of time (Knight, Wise, et al., 2017).

The process of learning can provide insights into understanding the nature of learning. Temporal analytics, within the field of learning analytics, is dedicated to exploring the learning process and its temporality. (Bogarín et al., 2018; Chen et al., 2016a; Knight, Wise, et al., 2017). Due to the temporal nature of learning, it is required to use proper techniques to capture the temporality aspect. There are three main benefits of utilizing temporal techniques for education and learning practices (Knight, Wise, et al., 2017; Reimann et al., 2014). Firstly, temporal analysis provides nuanced ways to explore data. Many researchers utilize these techniques to identify further temporal patterns that would otherwise remain unknown. Secondly, this analysis offers new methods for investigating theories and contributing to the development of these theories. This benefit yields two main outcomes: 1) it can solidify learning theories through further investigation and examination; 2) it can highlight gaps in theories and stimulate discussions about advancing theories in the educational science community. Thirdly, temporal analysis has introduced many techniques not previously utilized in the field. Exploring and incorporating nuanced techniques can contribute to developing new systems (e.g., automated tutoring systems) or improving their performance, ultimately aiding educational practices. Overall, the benefits of temporal analytics suggest that considering the temporal nature of learning in research is essential (Knight, Wise, et al., 2017).

However, the temporality aspect has often been neglected in applied learning research (Bogarín et al., 2018; Knight, Wise, et al., 2017). As Reimann posited, researchers often overlook the full potential of available information regarding temporality (Reimann, 2009). He stated that human learning is inherently cumulative, and research on temporality should consider both quantitative aspects (e.g., duration, transitions) and order. Therefore, obtaining an appropriate methodological approach to exploit available temporal information is imperative. This issue inspired my thesis to devote its major attention to exploring the affordances of techniques that can identify patterns in the learning process.

1.2. Motivation

Temporal analytics, a sub-area of the field of Learning Analytics (LA), is dedicated to bringing insights into learning concepts that would remain unknown without temporal analysis and its techniques (Knight, Wise, et al., 2017; Reimann et al., 2014). For instance, a study conducted by Kinnebrew and colleagues (2014) assessed the impact of feedback on the learning process (at the levels of cognitive and metacognitive) during learning engagement among middle school students. Despite insignificant results from the correlational test, this study highlighted the power of exploratory studies in understanding different aspects of students' learning behavior and relating them to knowledge building over time (Kinnebrew et al., 2014).

There have been many techniques used for temporal exploration of data such as visualization analysis (Riel et al., 2018a), frequent sequential analysis (Jovanović et al., 2017a), transitional analysis (Mahzoon, Maher, Eltayeby, & Dou, 2018a), network analysis (Kinnebrew et al., 2014), fuzzy mining technique (Beheshitha et al., 2015), and other techniques (Bogarín et al., 2018). Although we know about the differences between the techniques, it is not clear which type of questions they are most suitable to address, or which type of data they require (Knight, Wise, et al., 2017; Molenaar, 2014a). For instance, a comparison study conducted by Matcha et al. (2019) on the result from three prominent temporal analysis approaches in the detection of learning tactics and strategies, in a MOOC setting (Matcha, Gašević, Ahmad Uzir, et al., 2019) showed that different techniques can yield different results and interpretations for obtained learning strategies. Another comparative study was conducted by Chen and colleagues (2017) to explore two prominent sequential mining models including Lag-sequential Analysis (LsA) and Frequent Sequence Mining (FSM) (Chen et al., 2017a). The techniques provided different but complementary analyses on temporal patterns. These studies showed that there is no coherent guide to selecting certain temporal techniques to lead to achieving results that reliably uncover underlying phenomena.

Similarly, Knight et al. highlighted challenges in researching the temporal nature of learning (Knight, Wise, et al., 2017). These challenges are associated with 1) theoretical challenges around the concepts of the learning process, and 2) methodological and practical challenges in exploiting temporal data. The first challenge involves conceptual difficulties, particularly with the notion of temporality, which includes the concepts of elapsed time. The study emphasizes that these time concepts should inherently offer insights into the progression of learning, necessitating detailed measurement of the flow of time and sequence of events. The second challenge (methodological) arises from the quantity and complexity of temporal data, posing risks to the validity and generalizability of research findings. This is especially true when methodologies predominantly depend on traditional analytical approaches like correlational studies. Consequently, more sophisticated analytical methods are essential to evaluate time windows and analyze time units, aiming to accurately measure the progression of activities associated with learning. Moreover, the authors underlined the importance for researchers to be vigilant in ensuring the proper collection, transformation, and storage of temporal data, to render temporal analytics feasible.

Therefore, researchers need effective methods to manage and analyze temporal data. The methodological techniques that are not sufficiently capable of handling the data can compromise the validity of the research. The main contribution of this thesis is to provide guidelines that can guide temporal research to select effective methodological approaches to address their research questions.

1.3. Contributions

This thesis makes several contributions to the expanding field of temporal analytics, and thus, to learning analytics. In the next chapter, I delve into the connections between learning analytics and temporal analytics. These fields, having originated from the expanding use of technology in education, have influenced educational research and practices in recent years. In Chapter 2, I provide a detailed review of the key components of learning analytics and associated literature.

Chapter 3 proposes research questions (RQs) for this thesis. It also highlights the challenges in conducting temporal studies in LA and discusses the necessity of having a framework as a guideline to address these challenges.

Chapter 4 provides a systematic mapping study that enhances our understanding of the question types and methodologies employed in temporal educational research. The goal of this study is to dissect various elements of published research and investigate current trends in educational studies that specifically address the temporal aspects. Initially, I provide a detailed review of prior mapping research and associated guidelines to ensure the validity of this study. Then, leveraging a thematic coding method, I elaborate on patterns in temporal research components, including research questions being asked, data obtained at various granular levels, analytical techniques being utilized, and derived insights about learning.

In Chapter 5, I present the findings in the form of a framework to conduct temporal studies. This framework plays a crucial role in conducting two follow-up temporal studies in the next two chapters.

Chapter 6 describes the first case study to showcase and evaluate the proposed framework. It follows the provided guidelines to reveal the temporal nature inherent in students' self-regulated learning (SRL) behaviors. I elucidate how these guidelines assist in unveiling various temporal aspects related to SRL behaviors. This study discovered two facets of temporality. The first facet relates to the sequence of SRL phases. The study identified four categories of SRL processes based on phase transitions and the recurring nature of SRL. These SRL processes were then aligned with the types of iterative behaviors across SRL phases. These behaviors correspond to the theoretical self-regulating actions of students possessing different levels of SRL skills. We also statistically tested the connection between SRL processes and assignment grades. The study then shifts its focus to the second facet of temporality, relating to instances of time. It uncovers the temporal dynamics of SRL phase transitions by analyzing time profiles for shifts between SRL processes.

Chapter 7 adheres to the structure established in Chapter 6 as it presents the second case study. It utilizes the information process proposed in our framework to investigate temporality in students' posting-related behaviors. The research community has recognized the importance of personalized feedback based on students' needs (Bienkowski et al., 2014; Matcha et al., 2020; Vieira et al., 2018). To achieve this goal, we need to deepen our understanding of students' characteristics that should be used for personalization, in order to know what information to communicate to the student, and how to frame and present it to make the dashboards more effective in motivating students and leading to desirable behavioral changes. In this chapter, I elucidate and evaluate how our framework guided each step in conducting the study, leading to the discovery of the impact of visualization feedback on the dynamics of engagement within online discussion activities. Indeed, this study leverages diverse aspects of temporality to comprehend how the visualization feedback stimulates students' participation in discussion posts and how the visualization influences their re-engagement with discussions.

In the next chapter, I discuss how the framework differs from existing frameworks. Then, I reflect on our framework and its contributions to our case study. The insights gained from these studies not only provide practical implications for designing more effective SRL feedback systems (first case study) or online discussion dashboards (second case study) but also reinforce the value and adaptability of our proposed framework in different learning contexts. I also discuss the limitations and shortcomings of the framework in its application based on our case studies, highlighting areas where improvements or adjustments might be needed for future research.

The final chapter provides a summary of this thesis's contributions and explores potential extensions of its findings in future research.

In summary, this thesis advances the field of temporal analytics by mapping the current state of the field and providing a comprehensive framework for conducting temporal studies. At first, the mapping study aimed at broadening our understanding of trends in temporal educational studies. It illustrates the associations between the research questions asked by researchers, the data utilized at various granularities, and the analytical techniques employed, all while considering insights gained from learning. This evolution of the field provides an additional layer for reviewing studies that discuss aspects of temporality in educational research (Gašević et al., 2017; Knight, Wise, et al., 2017; Reimann et al., 2014). Furthermore, I propose a framework based on the mapping study's outcomes, and I conduct two follow-up studies to demonstrate how to take advantage of the framework to further explore the temporality aspects of learning. The two case studies, while serving as showcases and evaluations of the framework, made their unique contributions to their respective subfields of learning, thereby offering tangible implications for real-world educational contexts.

Chapter 2. Literature Review

This chapter explores the field of learning analytics as an emerging field of research by conceptualizing different components of learning analytics. We first demonstrate the importance of a framework and the components that constitute this interdisciplinary field of study. Following this, we discuss the significance of temporal research in learning analytics, emphasizing its profound relationship with the main field. Next, we explore how temporal research utilizes various data and techniques to discover insights about learning.

2.1. Structure and Key Components of Learning Analytics

The field of Learning Analytics (LA) is defined as a multidisciplinary field that focuses on optimizing learning through the collection, measurement, analysis, and interpretation of data about learners and their environments (Ochoa et al., 2017). Gašević et al. (2017) described the LA field as a multidisciplinary field that integrates various elements to improve learning practices, research, and decision-making (Gašević et al., 2017). They elaborated on three main dimensions of the LA field, including theory, design, and data science (Figure 2-1). These three dimensions are essential for the LA community due to the interdisciplinary nature of the field.

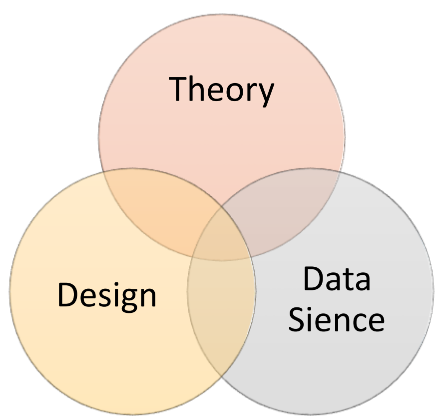

Figure 2-1. Three component of Learning Analytics field and it is sub-component.

The first dimension encompasses the theoretical foundation, which spans various disciplines including education, psychology, and sociology. This foundation provides a critical backbone for research practices and applications in LA. Imperfections in LA practices, where the theoretical foundation is not adequately considered, can hinder the interpretation of whether specific learning processes are engaged. Additionally, this theoretical foundation assists in explaining any inconsistencies across outcomes from different research studies.

The second dimension focuses on design, comprising interactive visualization, learning design, and study design. Interactive visualization, which influences learning outcomes, should be grounded in existing theories. Learning design, aimed at enhancing effective learning experiences, is rooted in theoretical foundations and impacts the interpretation of results from predictive models. Study design involves conducting research and assessing results based on established principles and theories.

The third dimension is data science, covering practices related to data collection, measurement, analysis, and reporting. A significant sub-area within this is feature engineering, which identifies indicators of learning processes, outcomes, and other activities from a diverse range of data. It is important that data science methods are integrated with a theoretical underpinning to yield meaningful insights.

In another review study on Learning Analytics (LA) studies, the authors categorized types of studies differently. They proposed five distinct categories: prediction models, learning theory, designed frameworks, applications, and data-driven decision-making (Hantoobi et al., 2021). They posited that LA studies comprise three main components: collected data, employed analytical techniques, and inferred insights leading to actions that impact the learning experience (Figure 2-2). Utilizing these components, the study proposed five research categories in LA.

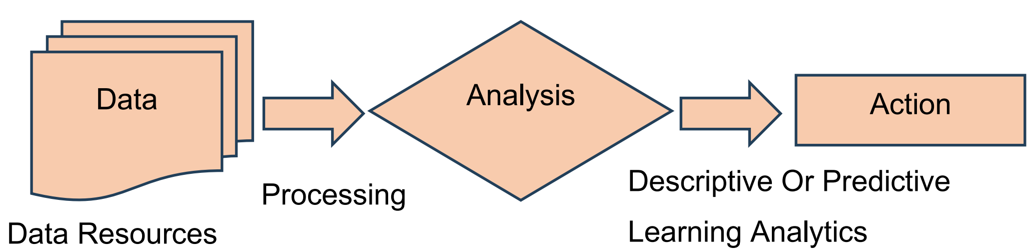

Figure 2-2. Learning Analytics (LA) process by Hantoobi et al., (2021).

In the first category, prediction models, the authors argued that predicting academic achievements requires considering diverse factors, such as platform interactions and feedback engagement. They emphasized that relying solely on traditional grades is insufficient; a multifaceted approach using various data points is essential to capture information about the learner's process. The second category highlighted the associations between theory and insights into learning. The third category discussed the importance of frameworks in conducting LA studies. A clear framework enables better curriculum design and understanding of educational outcomes. They discussed different frameworks, including those proposed by Gašević et al., (2017) and Greller & Drachsler, (2012), which focus on data quality and ethical considerations. Similarly, in the last two categories, the authors emphasized the importance of a framework to ensure the practicality in learning setting. They also highlighted the need for guidelines that consider data and take a holistic approach to understand and apply learning analytics.

All in all, the aforementioned studies emphasize the collection of data, analytical techniques used, and insights gained as the three main components of LA studies. Before delving into these components, I will discuss the importance of temporal analytics in LA and continue the rest of this thesis through a temporal analytics lens.

2.2. The importance of Temporal Analytics in LA

Knight and colleagues (2017), in their paper 'Time for Change: Why Learning Analytics Needs Temporal Analysis', argue that the essence of LA is deeply rooted in temporal data, which reflects actions indicative of learning activities. They highlight the necessity of investigating the temporal aspects of data, advocating for more in-depth research within this sub-field of LA (Knight, Wise, et al., 2017). Therefore, temporal analytics, as a sub-field of LA, is dedicated to exploring the learning process and its temporality, thereby offering valuable insights into the learning process (Bogarín et al., 2018; Chen et al., 2016b; Knight, Wise, et al., 2017). Temporal analytics holds numerous advantages for education and learning practices, providing nuanced approaches to data exploration (Knight, Wise, et al., 2017; Reimann et al., 2014). Researchers widely utilize these techniques to uncover temporal patterns that would otherwise remain undetected without temporal analysis. Thus, temporal analytics significantly contributes to the LA field, adhering to the belief that “Let's not forget: Learning Analytics are about learning” (Gaševic et al., 2015).

2.2.1. Types of Data in Educational Research

At the outset of designing a study in the field of LA, it is crucial to identify the available data and determine the types of temporalities for analysis. Nistor and Hernández-García (2018) highlight that analyzing a variety of data sources can deepen our understanding of educational dynamics, thereby improving our ability to predict outcomes and implement effective interventions (Nistor & Hernández-Garcíac, 2018a). In their review, they categorize data types as follows: log data, self-reported data, eye-tracking data, online dialogue data, and visual learning data. Log data, often used at both individual and collaborative levels, is the most prevalent. This type of data, also referred to as 'event-stream data' in other studies, typically consists of clickstream information generated during learner interactions with educational materials. For example, log data from an online course, derived from learners' clickstreams, captures records of their information access on online resources, including the timing and sequence of these interactions (Siemens & Baker, 2012). Self-reported data, collected directly from learners through methods like surveys or diaries, represents another significant data type. Eye-tracking data, capturing where learners focus their gaze on a screen, is frequently used in studies analyzing visualization feedback or user interface design. Online dialogue data records textual interactions among learners, instructors, or assistant tutors, providing rich insights into knowledge building and associated sentiments. Lastly, visual learning data in LA, applicable with the Internet-of-Things technologies, includes data from wearable devices.

Arita Liu (2023) investigated the interaction pattern in asynchronous online discussion (AOD) to understand how students learn through discourse by examining timing, social aspects, and the discussions involved (Liu, 2023). The author utilized three types of data: log data, discussion content, and achievement data. Utilizing these data, Liu identified that students' participation patterns (temporal patterns) were associated with different discussion formats, achievement levels, and students' group configurations in the context. By analyzing temporal patterns, Liu was able to discover the importance of three factors in successful asynchronous online discussion design: encouraging time management, considering student motivation, and allowing autonomy in discussion.

These insights could guide the development of effective instructional strategies and grading rubrics in the future.

From the perspective of temporal analytics, accurately recording the timing and sequence of events in data is crucial. This is particularly important in analyzing event-stream data from log records, where the order of events can reveal aspects of the learning process. For instance, a study by Sher et al., (2019) delved into multiple modalities within log data to examine students' engagement in online discussions. In this context, 'modalities' referred to the use of various devices by students. The researchers created profiles based on the sequence in which students switched between these devices, termed 'modality sequence profiles'. These profiles were then correlated with measures of online engagement and academic performance. The study underscored that the temporal patterns in the usage of different modalities (such as cellphones or laptops) are linked to students' overall academic achievement.

In other studies, different types of data were collected and engineered to reveal temporality in the learners' behavior. For instance, studies combined eye tracking and log data to explore the temporality in interactions of learners (Chiou et al., 2019; Jin & Yu, 2019; Tsai et al., 2012). These studies found that temporal patterns in eye movements are associated with the participants content awareness. Other research used emotional data and log data to understand temporal engagement in online discussions (Liu et al., 2019; Lund et al., 2017), uncovering the indicator of emotions that can be preceded by key learning moments. Some studies collected gaming interaction data to assess learning impacts (Minović & Milovanović, 2013; Taub & Azevedo, 2018; Yang & Lu, 2021), revealing that specific temporal patterns in gamification learning were associated with improved learning performance. Additionally, other studies have generated novel features that represent the learners' temporal behaviors (Caprotti, 2017; Hansen et al., 2017; Lum et al., 2013; Lwande et al., 2021), identifying features like session duration and time-of-day as significant predictors of learning success. These varieties in data types pose a methodological challenge to temporal studies: selecting the right technique for the data at hand. Another complication arises from the granularity of data, which is often unclear, making it challenging to identify an appropriate technique suitable for specific data to derive meaningful insights.

2.3. Techniques in Temporal Analytics

The huge amount of data from the learner requires proper techniques for effective analysis. Within the domain of temporal analytics, a significant portion of the analytical techniques are derived from the data science discipline, prominently featuring statistical analysis and machine learning as central pillars. Temporal analytics not only relies on these standard methods but also incorporates others like Network Analysis, which studies the interconnections within various states; Time-Series Analysis, focusing on changes in data points over time; Frequent Sequence Analysis for identifying patterns or sequences in data; and Markov-Chain Analysis, employed for assessing the next events based on current data. These process-centric techniques focus on analyzing the sequences and flows of processes over time, enhancing our understanding of the temporal nature of learning. These techniques have recently received attention in the Learning Analytics field, shedding light on the dynamic nature of learning. Furthermore, comparative studies have been conducted to illustrate the affordance of these techniques in revealing different aspects of temporality in learning.

One prominent example is a comparative study conducted by Matcha et al., (2019) on three popular temporal analytics techniques in the detection of learning tactics and strategies, in a MOOC setting. Learning strategies are defined as follows:

“Any thoughts, behaviors, beliefs or emotions that facilitate the acquisition, understanding, or later transfer of new knowledge and skills”. Learning strategies define how students use different learning tactics. Patterns in tactics can show characteristics of an individual’s learning. This study collected data from students in Python programming courses and coded based on activities such as reading materials, taking quizzes, etc. Due to large, diverse, complex data from learning platforms (especially MOOC), it has been a challenging task to analyze the data, researchers have adopted varied techniques to understand the tactics and strategies adopted by learners. This study was dedicated to investigating three prominent techniques in analytics including sequential analysis, process analysis, and network analysis. Three techniques are utilized to understand the obtained learning strategies (Figure 2-3 shows the pipeline for the methodology).

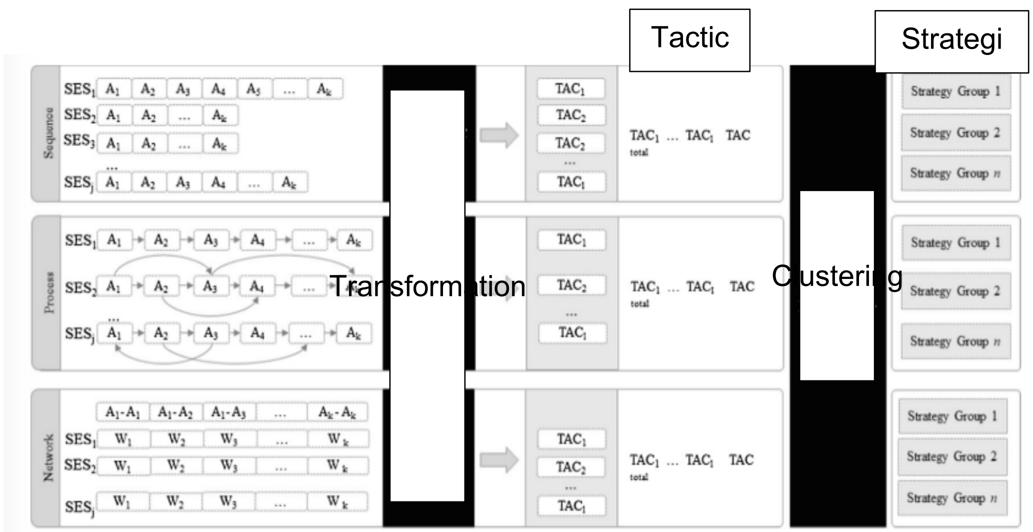

Figure 2-3. Pipeline of different techniques to reveal learning tactics by Matcha et al., (2019).

In Matcha and colleagues' study, the first technique is network analysis, which considers the connections of learners and identifies strategies and tactics by constructing a network based on the frequencies of their occurrences. Utilizing epistemic network analysis, the study generates a network of co-occurrences of activities. Subsequently, the Ward algorithm clusters the network to produce learning tactics. The second technique is sequential analysis, emphasizing the sequence of actions associated with learning. This study computes the transition frequencies and the similarity in the sequence of actions in each given learning session. Then, using the

Ward clustering algorithm, it groups sequences and proposes tactics. The third technique is process analysis, or process-oriented data analysis, focusing on the timing of event occurrences. This technique provides the probability of transitions in defined time intervals. Subsequently, the study uses a clustering algorithm to cluster processes.

The results showed that diverse strategies in Network Analysis and Process Analysis yielded similar patterns. For example, the most frequent action was practicing exams followed by taking quizzes. However, the Sequence Analysis technique captured some less similar but more diverse patterns. In general, the three approaches produced similar results, to some extent, in detecting strategies. However, in terms of learning strategies groups (student groups), it was observed that Network Analysis produced different strategies in comparison with Process and Sequential Analysis. All strategies can be interpreted and explored in accordance with theory. This study did not claim the superiority of one approach over another. It demonstrated that different approaches created different data representations, which were fed into a clustering algorithm to detect tactics. The choice of the clustering algorithm is also crucial in handling specific data structures.

This study showed that selecting different temporal analytics techniques could yield different results. Each approach can reveal a different dimension of the data, which can be explored and investigated for its alignment with theory. Therefore, we conclude that there is room for developing general guidance for using different techniques in the temporal analytics field.

Similarly, another comparative study investigated two prevalent sequential mining techniques: Lag-sequential Analysis (LsA) and Frequent Sequence Mining (FSM) (Chen et al., 2017b). These techniques were employed to detect productive threads in knowledge-building discourse in a collaborative setting among students. To quantify online behavior, the study measured activities such as the number of posts, words, log-ins, etc. Furthermore, these measures should be combined with social aspects, such as the role of instructor intervention, to capture the social dynamics of groups, which can be an indicator of online engagement. Therefore, the study aimed to unveil the temporality dimension of online engagement by focusing on the transition between activities, such as moving between notes (different types of notes) or inserting new threads. Using LsA, the study examined immediate transitions (lag = 1), and the results showed that productive threads have a significantly higher number of transitions, suggesting that students engage more with resources in such threads. Furthermore, the study assessed indirect transitions (setting lag = 2), revealing that productive threads also have more frequent indirect transitions. On the other hand, FSM can reveal more frequent sequences than the immediate occurrence of sequences. In fact, applying FSM analysis can uncover the contingencies among productive sequences. The transitions between sequences are more frequent when sequences represent problem-solving attempts, which are considered productive sequences. Overall, the study concluded that both techniques are complementary for the temporal mining problem: while LsA offers a pairwise comparison, FSM is effective in dealing with gaps among activities.

Furthermore, researchers have been consistently experimenting with new techniques on their data to elucidate the learning process. For instance, a study by Sharma et al. (2020) utilized a Hidden Markov Model (HMM) to assess the effort students put into tasks, a factor known to affect educational outcomes but difficult to observe directly. This research used diverse data types, such as eye-tracking, EEG, and facial expressions, to categorize the behavior patterns of 32 students during a self-assessment activity (Sharma et al., 2020). By employing Hidden Markov Models (HMMs) and the Viterbi algorithm, the study predicted the intensity of effort students would likely apply in future tasks based on these patterns. This approach also helps to identify the right moments to give feedback to students. As another example, a study explored timestamps of log files to understand the immersive level of the learners' experience in a virtual science world (Reilly & Dede, 2019). They utilized principal component analysis to create time-series trajectories, enabling them to track and group students' activities over time, and find different patterns and pathways taken during the learning process. These patterns were then linked to different learning outcomes, proposing how to provide effective support to various groups of learners. In other studies, various techniques have been utilized to explore the temporal nature of data such as visualization (Riel et al., 2018b), frequent sequence analysis (Jovanović et al., 2017a; Nazeri, Hatala, & Salehian Kia, 2023), transitional analysis (Mahzoon, Maher, Eltayeby, & Dou, 2018b), network analysis (Kinnebrew et al., 2014), fuzzy mining techniques (Beheshitha et al., 2015), and other techniques (Bogarín et al., 2018; Hatala et al., 2023a). Although we know about the technical differences between the analytical techniques, it is not clear which type of questions they are most suitable to address in the educational context, which type of applications they can furnish, or which type of data they require (Knight, Wise, et al., 2017; Molenaar, 2014b). This issue highlights the complexity of selecting an appropriate technique suited for the collected data.

Therefore, although the aforementioned studies aimed to illuminate the comparisons between these techniques, they are largely confined to the specific data that researchers possess. Consequently, a framework is needed to align various types of data with the appropriate temporal techniques.

2.4. Insight About Learning in Temporal Analytics

After collecting data and utilizing the technique, it is crucial to understand what types of insights about learning can be inferred through the analysis, which is the most important outcome in educational technology research, as Gaševic et al., (2015) asserted in their paper, “Let’s not forget: Learning analytics is all about learning”. These insights can be inferred from analyzing available data and often need to be aligned with theory. For instance, studies examining participation engagement in discussion forums can shed light on learning progression (Boroujeni et al., 2017; Hatala et al., 2023b; Huang et al., 2019). The indication of learning can be measured through theories such as communal knowledge building or communities of inquiry. The aim is to enhance learning progression by improving engagement.

Furthermore, as mentioned earlier regarding techniques in this chapter, using different analytical techniques can yield various insights about learning. Currently, there is a gap in studies, as it is unclear which technique, paired with available data, leads to specific insights about learning. Therefore, this thesis aims not only to explore the types of learning insights revealed through temporal educational research but also to establish the connection between insights, data, and temporal techniques. By identifying these insights, the study aims to improve our understanding of how temporal educational research can contribute to comprehending learning processes and outcomes, making the study more impactful. As emphasized in Learning Analytics, when conducting a temporal study, it is vital to always ask, “What impact(s) will this study create on learning?” (Knight et al., 2017).

From a temporal perspective, studies have been applying analytical techniques to their data to gain insights into the dynamic nature of learning. For example, the variety in students' sequential engagement with learning materials – such as watching videos, taking quizzes, and reading resources – can be identified as an indication of learning. Such insights often represent the behavioral aspect of how learning happens (Fan et al., 2021a; Hatala et al., 2023b; Jovanović et al., 2017a), while some studies aim to understand dynamics within asynchronous discussion forums to comprehend communal knowledge building. Other studies focus on predicting student behavior (Lwande et al., 2021; Scherer et al., 2012; Shin et al., 2021), while others prioritize forecasting academic performance (Hu et al., 2014; Umer et al., 2019; Van Goidsenhoven et al., 2020). However, these predictions sometimes miss deeper insights into the actual learning process. A challenge in this domain is determining the proper combination of data and analytical techniques to yield specific types of insights. Hence, this thesis aims to bridge this gap by illustrating the associations among the types of data collected, techniques employed, and insights gained about learning within temporal studies.

Chapter 3. Research Questions

In this chapter, I discuss the challenges associated with conducting temporal studies. With the growing interest in temporal studies, it is crucial to have clear guidelines to frame and conduct such research. To fill this gap, I introduce a framework and provide two real-world examples that both showcase and evaluate its effectiveness. In doing so, I formulated four research questions to guide this thesis.

3.1. Issues in Conducting Temporal Study

As already pointed out, temporal analytics play an important role in understanding learners' behavior and potential outcomes. However, like any intricate field, temporal studies also come with inherent challenges.

1. **Diverse Temporal Data Types**: To conduct a temporal study, it is important to note that the data available for analysis determine what types of temporalities would be considered for analysis. Different categorizations have been proposed for data, such as event stream data, self-reported data, eye-tracking data, and others (Nistor & Hernández-Garcíac, 2018b). However, this classification of data types does not consider the granularity of data. Furthermore, some studies have gathered data from varying timeframes, such as 50-minute class sessions, while others have focused only on the sequences of learning events, such as users' logins, readings, and posting discussions (Chen et al., 2018a; Knight, Wise, et al., 2017). This diversity can complicate the process of comparing data across different studies. Therefore, this thesis delineates classes of data that have been utilized in temporal studies with respect to their granularity. Such categorization can guide researchers in selecting the most fitting analytical techniques suited for their data.

2. **Methodological Complexities in Temporal Analytics**: To model temporality and reveal patterns in the data, it is important to implement appropriate temporal analytical techniques. Several temporal techniques have proven capable of modeling the temporality aspect of learning. Examples include obtaining frequent sequence analysis to model the block of actions of students and assessing the differences between these blocks (e.g., Chen et al., 2017c; Jovanović et al.,

2017a), assessing the transition between actions to model the likelihood of the consequent action (e.g., Jo et al., 2014; S. Y. Wu & Wang, 2020), using network analysis to model the connection of actions of students to measure cognitive and meta-cognitive processes of students (Kinnebrew et al., 2014), and other techniques to model temporal concepts in learning (e.g., Bogarín et al., 2018). However, the trend in the application of these techniques in temporal studies remains unclear. Identifying such trends could significantly enhance our understanding of the affordances provided by these techniques, based on the data collected. Therefore, this thesis also aims to review techniques utilized in recent years and discuss their associations with the collected data.

3. **Moving from Theory to Actionable Insights**: Despite the growing focus on the temporal aspects of learning, there remains a lack of methods that yield actionable insights. The challenge lies in bridging the gap between theoretical constructs and practical applications (Chen, Wise, Knight, & Cheng, 2016). This thesis showcases how the relationships between research questions, data collection, and analytical techniques lead to insights about learning, which are essential for actionable practice.

3.2. The Imperative of a Framework for Temporal Analytics

Temporal research has received considerable attention in recent years (Fan & Saint, 2021a; Gašević et al., 2017; Knight, Wise, et al., 2017; Lee, 2021; Reimann et al., 2014; Saint et al., 2021). With the increasing focus on the temporal aspects of learning as an emerging area of study in educational research (Chen et al., 2018), there is a new need for a guiding framework to help researchers navigate this area. Such a framework should provide clarity, mitigate redundancy, and offer a systematic way to approach and handle the challenges of temporal analytics.

To create a framework, we propose utilizing a data-driven approach based on a substantial body of real-world research for drawing conclusions based on the “Method Framework for Design Science Research” (Johannesson & Perjons, 2021). In doing so, a systematic mapping study serves as a foundational step in providing the data, and thus, insights needed to establish the framework. Such a study is vital for consolidating existing knowledge and offering a comprehensive overview of current research

(Petersen et al., 2008). Additionally, it can uncover recurring themes and patterns, highlighting various data types used in temporal studies, such as event stream, self-reported, and eye-tracking data (Nistor & Hernández-Garcíac, 2018). However, previous research has often neglected the granularity of these data types, an aspect our mapping study seeks to address. We aim to categorize different data types and their respective granularities. Additionally, the interdisciplinary nature of temporal analytics, inherited from the LA field, poses a methodological challenge in selecting analytical techniques that can yield deeper insights into learning processes (Chen et al., 2018). Our mapping study will also address this challenge by identifying and understanding the analytical techniques currently in use. This approach sets the foundation for future standardization and methodological guidance in the field. With this goal in mind, the first two research questions of this thesis are dedicated to constructing this framework.

Upon proposing the framework based on the findings from the mapping study, the next step is its practical implementation. Therefore, we design and conduct two follow-up case studies for two reasons:

1. Evaluate the framework in real-world scenarios to ensure the applicability and reliability of the framework. This step involves applying the framework to actual data and scenarios in the field, observing how it performs under diverse conditions, and suggesting necessary adjustments based on these observations.

2. To demonstrate the practical use of the framework, thereby serving as a model for researchers in temporal analytics and Learning Analytics (LA). This involves showcasing how the framework can be applied in specific temporal research, illustrating its utility in simplifying complex temporal data analysis, and providing clear, step-by-step examples of the framework in action. This practical demonstration aims to guide researchers on how they can adopt the framework in their own studies, highlighting its versatility and effectiveness in various research contexts.

3.3. Research Questions

The motivations and background outlined in the previous chapters led me to formulate the following research questions that guided this thesis:

RQ1: In the area of temporal analytics in educational studies, (a) What types of research questions have been asked? (b) What types of data have been obtained and engineered? (c) What analytical techniques have been utilized in temporal studies? (d) What types of insights about learning have been discovered?

The contribution of the first research question is to aid in understanding how temporal educational research is conducted and the insights it can provide. I analyze different components of existing studies and explore current trends in educational research that explicitly consider the temporal aspect. The thesis focuses on the research questions that have been answered, the data collected, the features engineered, and the types of analytical techniques utilized in published temporal educational research. Another contribution is to explore the types of insights about learning that have been uncovered through temporal educational research. By identifying these insights, the thesis provides a better understanding of how temporal educational research can contribute to our understanding of learning processes and outcomes. Overall, RQ1 provides a comprehensive overview of the current state of temporal educational research and its contributions to the field of education.

To address this research question, I conducted a systematic mapping study to identify and categorize different components of educational temporal studies.

RQ2: What are the associations between research questions asked in the existing literature, data being collected, analytical techniques utilized, and insights about learning discovered?

After identifying different aspects of temporal studies in RQ1, RQ2 aims to explore the associations between those aspects in existing research. This exploration provides valuable insights into which techniques are most suitable for different types of research questions and available data, which will be beneficial for researchers conducting future temporal educational research. To elucidate these associations, we analyzed a corpus of published research to uncover the relationships between the questions asked, data collected, and the techniques employed in these studies. Furthermore, I segmented the associations based on distinct learning insights derived from this research.

RQ3: Based on the evidence uncovered, what framework can be developed to guide the selection of temporal techniques and learner data for deeper insights into the learning perspective?

Building upon the findings from RQ1 and RQ2, RQ3 focuses on developing a framework informed by existing evidence. To address this research question, we discuss the key components of the temporal model in educational studies. Selecting the correct component is crucial for optimizing insight into patterns of user behavior. Owen & Baker (2020) introduced a framework composed of three main components: data collection, feature engineering, and data analysis, initially designed for behavioral analysis in serious game design. Their methodology allows for systematic extraction of detailed features about user interactions and behaviors, emphasizing that synergy between data collection and analytical techniques enhances the analysis.

I adapted the core structure of Owen & Baker's framework to show how the thesis findings (from prior RQs) can guide researchers in data collection, feature engineering, and analysis for a temporal model.

This framework, distinct from a general research approach, offers a structured series of steps specifically tailored for temporal analytics in educational research. Unlike the broader research process, which broadly outlines steps from formulating research questions to analyzing results, this framework provides a focused, operational guide. It also includes a web tool: a curated database of research exemplars demonstrating techniques and data usage for gaining temporal insights. This specificity is vital in navigating the complexities of temporal analysis, where many data types and analytical techniques can be combined for varied research objectives. The framework thus serves as a guide, steering researchers toward selection of data and techniques, thereby enhancing the potential to derive meaningful insights from temporal dimensions in educational contexts.

RQ4: How can studies in the educational domain utilize our framework as a guideline for conducting temporal analyses?

This thesis presents two follow-up case studies as both showcases and evaluations to illustrate how the proposed framework effectively captures various temporal dimensions of learning and guides researchers in their research study design and execution. In the first case study, we adhere to the information process outlined in the framework to reveal the temporality of Self-Regulated Learning (SRL) behaviors in students. We identify studies with similarities in data and research questions to unveil indicators of learning as insights into the learning process. With these references and our framework, we employ various techniques to capture the multifaceted temporal aspects of SRL engagement behavior. The outcome offers practical suggestions for educational providers, guiding them on the optimal times for intervention to enhance SRL engagement levels. In the second case study, we follow the information process proposed in the framework to explore the temporality in posting-related behaviors with respect to visualization feedback in asynchronous online discussions. This case study demonstrates how the framework guides the selection of temporal techniques appropriate for the data collected, thereby revealing students' temporal behavior in response to visualization feedback and their subsequent posting activity. By applying the framework, we are able to systematically analyze and understand how visualization feedback influences students' engagement in these discussions. The insights gained from this study not only validate the framework's utility in practical research scenarios but also contribute valuable knowledge on student engagement patterns in online learning environments.

Chapter 4. Systematic Mapping Study

This chapter is one of the main contributions of this thesis. It offers a structured overview of the relevant research area and highlights current trends. The results provide a comprehensive perspective, which is essential for proposing a framework to guide researchers in temporal studies within the educational domain, thereby answering the first two research questions. Before addressing the RQs, we provide a methodology to conduct a systematic mapping study and elaborate on the key steps to conduct the study. In the results section, we break down RQ1 into four components, each addressed individually. For our RQ2, we use two separate sections. Initially, we illustrate the associations between the research questions asked, the analytical techniques utilized, and the insights gained about learning. Because data granularity is crucial, we dedicate a separate section to highlight how the data, with its specific granularity, is associated with the analytical methods and the insights obtained. In the discussion section, we discuss our findings and highlight their impact on the field of learning analytics, underlining key trends and takeaways. Finally, the limitations of the study are provided.

Most of this chapter is derived from previously published work as cited below:

Nazeri, S., Hatala, M., & Neustaedter, C. (2023). Associations of Research Questions, Analytical Techniques, and Learning Insight in Temporal Educational Research: A Systematic Mapping Study. Journal of Learning Analytics, 10(2), 68–84. https://doi.org/10.18608/jla.2023.7745.

Additionally, we provide complementary information about the data used, as presented in section 4.2.2, 4.2.6, and 4.3. Furthermore, this chapter addresses the first two research questions.

4.1. Method for Conducting a Mapping Study

Kitchenham et al. posited that a systematic mapping study provides a wide literature review to demonstrate the quantity and structure of evidence for decision-making (Kitchenham et al., 2011). In other words, the main goal of systematic mapping study is to understand the nature of existing work by exploring the trend in research area. The main difference between a systematic mapping study and a literature review is that, in a mapping study, the aim is to provide classification and structure of the research area. In a systematic literature review, the aim is to synthesize evidence to address certain research questions (Petersen et al., 2015). According to Petersen et al., the main steps to establishing a systematic mapping study in learning analytics are: 1) defining the protocol for the mapping study, 2) conducting data collection process, and 3) analyzing and summarizing the data (Petersen et al., 2008). Figure 4-1 shows the process of establishing a systematic mapping study.

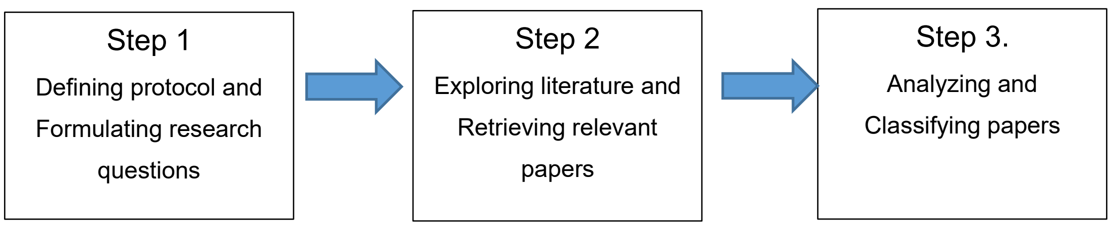

Figure 4-1. The process of conducting a mapping study

Mohabbati et al. (2013) adopted this framework and offered a comprehensive map of existing research and synthesizing current evidence on the integration of two paradigms: Service-Orientation (SO) and Software Product Line Engineering (SPLE). The results of their mapping study highlighted research challenges and provided direction for future research (Mohabbati et al., 2013). Similarly, Kitchenham et al., (2011) utilized the Petersen guidelines to emphasize the advantages of mapping studies as foundational baselines for advancing research in software engineering. Such baselines can serve as starting points for subsequent research endeavors (Kitchenham et al., 2011). Therefore, this thesis adopted similar approach to advance the field of temporal analytics and learning analytics by providing baseline framework for future temporal studies.

4.1.1. Step1: Defining Protocol and Formulating Research Questions

Defining a protocol in this study includes the following stages: identifying the data sources, describing the search and selection strategies, and describing the method for extracting and analyzing the studies. In this step, we are also proposing research questions.

Data Sources

To establish an exploratory search, we used digital libraries and searched through journals, conferences, and workshop proceedings in the area of education and educational technology from 2017 to 2021. We have chosen December 31, 2021 as the end date for the full completed calendar year and performed yearly searches (see below) backward, until we have reached the number of papers that we could feasibly examine within the timeframe and resources available for this study, i.e January 1st 2017. Coincidently, by 2017, the temporal analysis in learning analytics attracted the level of attention to warrant the call for a special issue of the Journal of Learning Analytics, which appeared in late 2017 (Vol. 4 No 3) and early 2018 (Vol 5, No 1). Our digital search included searching digital libraries, including the search engine at our university library, ACM digital library, IEEE digital library, Science Direct, and Google Scholar. We also manually searched the publishers' websites for the top 10 publications listed in Google Scholar's venue rankings in the category of Educational Technology[1]. These venues are listed in Table 4-1. Although utilizing multiple search strategies yielded many duplicates, using varied sources helped us to overcome limitations of these sources to execute the complex queries.

Table 4-1. The list of venues searched manually via their webpage (Scholar, n.d.).

Rank	Publication
1.	Computers & Education
2.	British Journal of Educational Technology
3.	The Internet and Higher Education
4.	Journal of Educational Technology & Society
5.	Education and Information Technologies
6.	The International Review of Research in Open and Distributed Learning
7.	Educational Technology Research and Development
8.	Interactive Learning Environments
9.	Computer Assisted Language Learning
10.	International Journal of Educational Technology in Higher Education

[1] https://scholar.google.ca/citations?view_op=top_venues&hl=en&vq=soc_educationaltechnology

4.1.2. Step 2: Retrieving Papers

Our search strategy to identify keywords and construct search queries follows guidelines from (Dickersin et al., 1994).

Identifying Query Keywords

We designed the following stages to ensure our search strategy includes a variety of papers that cover the area of interest. Table 4-2 shows the identified search keywords in each stage.

1) **Identifying the general search keywords and terms** based on the study's research questions. Accordingly, our RQs generally focus on "temporal analytics" and "learning analytics".

2) **Finding more keywords and terms used in prominent studies** in the area of "temporal analytics" and "learning analytics". In this stage, we selected an editorial paper in the special issue of *Journal of learning analytics* (2018) that focuses on "*critical issues in designing and implementing temporal analytics*" (Chen et al., 2018a). The paper reviewed literature on temporal analytics, and we extracted the authors' keywords from the paper. Further keywords were also extracted from the papers within the special issue (Chen et al., 2017c, 2018a; Knight, Wise, et al., 2017; Mahzoon, Maher, Eltayeby, & Dou, 2018b; Riel et al., 2018b). As a result, we identified 55 different keywords, and we selected the top 20 of the most frequent and relevant to temporality.

3) **Identifying synonyms and alternatives**. To identify synonyms, we searched a different area of educational technology. For instance, temporal analysis is a commonly used term for the concept of time for analysis in the learning analytics field. However, there are some closely related terms to temporal analysis, and many authors used those terms interchangeably. For instance, the term educational process mining is widely used in the Educational Data Mining (EDM) field (Bogarín et al., 2018). It seems that, in EDM, process mining is analogous to temporal analysis in LA. In the field of behavioral

psychology, Bakeman used sequential analysis for the same purpose as temporal analysis (Bakeman & M.Gottman, 1997). Furthermore, the outcome from stage 2 helped us to identify more similar terms. In this stage, we arranged keywords into two subgroups: 1) keywords that imply learning and theory; 2) keywords that are associated with analytical techniques.

4) **Simplifying the keywords to comprehend relevant terms**. In this stage, we simplified the keywords to cover similar words that may not be covered in stage 3. In doing so, we used special characters such as an asterisk (*) to specify characters in the keywords that can vary without altering meaning. For example, *sequen** includes *sequence*, *sequential, or sequencing.* This format is supported by our targeted online databases. In the case of not supporting this format, we manually inserted all possible keywords.

Table 4-2. Extracted keywords to generate a search query.

Stage	keywords
1	temporal analytics, learning analytics
2	learning analytics, sequential analysis, temporal analytics, self-regulated learning, knowledge building, educational data mining, teaching method discourse, discussion, community of inquiries, frequent sequence mining sequence data mining, sequence data model, teaching method, tempora database, process analysis, lag analysis, process, interaction sequence predictive model, cluster analysis, context effect, explanatory power, hol grain
3	learning analytics, educational technology, educational data mining, temporal analytics, sequential analysis, process analysis, process minin sequential mining, lag analysis, knowledge building, interaction sequenc cluster analysis, predictive model
4	learning analy*, educat* tech*, sequen* analy*, temporal analy*, process analy*, lag analy*, cluster analy*, predic*, predic* model, educat* data mining, process mining, knowledge building, interaction sequen*

Generating Search Queries

Having the keywords, we used logical operation (AND/OR) to generate search queries (Table 4-3). We defined three types of queries, and used a combination of these queries to construct our search query:

i) A query that covers the general area of educational technology.

ii) A query for the specific area of temporal research. We aimed to cover the extracted keywords from literature in the previous stage, using AND/OR operations.

iii) Generating the main search query by combining previous queries.

Table 4-3. Search queries used to extract relevant papers

Type	Search query
i	"learning analy*" OR "educat* tech*" OR "educat* data mining" OR "teach*" OR "pedagog*"
ii	("temporal*" OR "sequen*" OR "process" OR "lag") AND (analy* OR "mining" OR "model" OR "cluster" OR "predic*")
iii	Query(i) AND Query(ii)

Inclusion and Exclusion Criteria

To select relevant studies that can address our research questions, we applied Search Query (iii) Table 4-3 in all search engines and venues over the last five years (from January 2017 to December 2021). To ensure the relevance of papers in our corpus, we excluded studies that did not focus on temporal aspects of learning. Firstly, we removed duplicate papers from different sources. Next, we carefully reviewed abstracts and selected studies that focus mainly on the temporal aspect of learning and eliminated papers without attention to temporality in their abstract. The last excluding stage encompasses scrutinizing papers and reviewing sections of articles. This stage was accomplished during the qualitative coding of papers (discussed in the next section). Our main aim in the mapping study is to organize the studies and the information within the studies. However, to include a paper, the selected paper should encompass clear objectives and methodology, as well as have a minimum description of the students' learning progression or temporality in the method. As a result, 176 articles were retained. The flow chart of the selected papers in each stage can be seen in Figure 4-2.

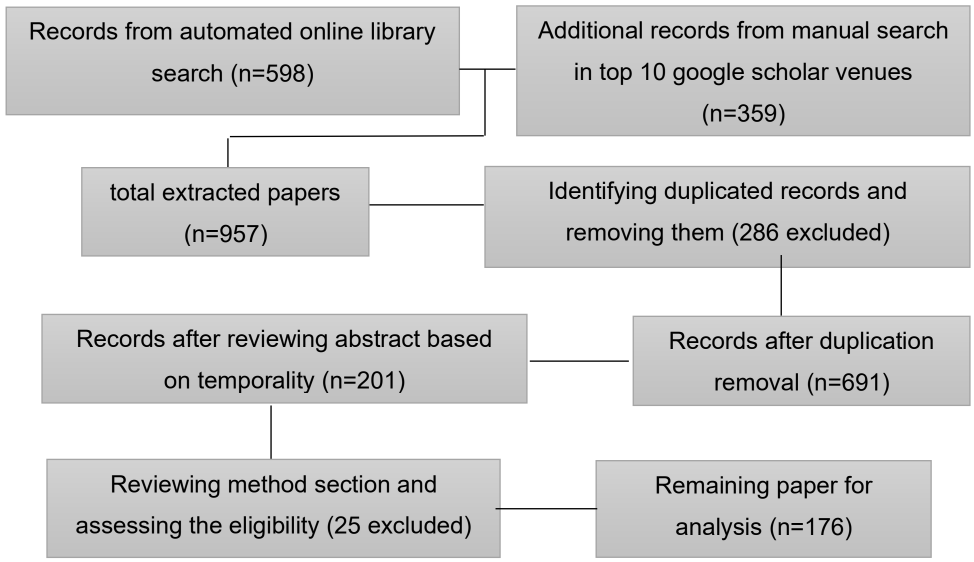

Figure 4-2. The number of selected papers in each stage.

4.1.3. Step 3: Developing Classification Scheme and Summarizing Results

The classification scheme is designed to characterize studies with respect to their research question focus, analytical technique, and obtained insight about learning. This study used thematic analysis to create a coding schema. The method has been widely used in qualitative research and term analysis (Basit, 2010). The thematic coding method is useful for coding descriptive terms in literature where the authors propose research questions, utilize the analytical technique, and discuss contributions and insights. At the higher level, the coding method helped us to identify the type of study and its contribution within each paper. Next, we were able to categorize different aspects of studies to address our research questions. To conduct a trustworthy thematic analysis, we followed Nowell and colleagues' guideline that provides a step by step approach including familiarizing ourselves with data, generating initial codes, searching for themes, reviewing themes, defining themes, and reporting (Nowell et al., 2017).

Identifying Papers' Sections for Coding

In the first step, to familiarize ourselves with the literature, we reviewed different sections of studies and identified those sections that matched our RQs. For example, to find out what *research questions* have been answered, we focused on the introduction and the study's research questions. Likewise, to code the *utilized analytical technique*, we reviewed the method section. To code the type of *inferences about learning*, the sections of results and discussions were reviewed. In cases where the paper did not follow a mainstream structure we searched for the pertinent information in other parts of the paper. The full list of the sections can be seen in Table 4-4.

Table 4-4. Coding sections were chosen for addressing our RQs.

Research coding sections	Description
Research questions / Research focus	Includes codes that show the main focus of the research. The code is obtained from the study explicitly mentioning the aims of the research.
Data Types	The type of collected and processed data that can be found in method section
Analytical techniques	The analytical techniques used in the research that can be found in method section.
Inferences about learning	The type of insights and inferences about learning that the paper offers. This can be mainly found in result and discussion sections.

Generating Initial Codes

In this study, we have worked with three themes that were directly mapped to the research questions, as listed in Table 4-4. In the second step, through the iterative process, we produced an initial set of codes for each section in Table 4-4, based on a detailed reading of the identified sections in 50 papers in our corpus. In this study, the first author conducted all the coding processes step by step, and the reliability of the produced codes was assessed by an expert iteratively. To ensure consistency of the codes, aside from the expert review, the data was revisited and recoded several times, as described below. As the papers in our corpus were typically coded with multiple codes in each theme, measures for interrater reliability were not used to measure the quality of the coding scheme. For full transparency, to support confirmability, Appendix A shows the assigned codes for all the papers.

Reviewing and Finalizing the Codes

After the initial codes from 50 papers were stabilized, a random sample of 10 papers was coded independently by the two authors, discrepancies were discussed, and the coding schema and code definitions were updated. The majority of adjustments in this phase involved determining the boundaries for the codes, i.e., how prominent the research question was, the analytical technique, and the level of theoretical grounding to support assigning the learning insight code. Another set of 30 papers was coded independently with the revised set of codes, and a final adjustment to the schema and code definitions were done. After discarding the codes assigned to papers in the development stage, the first author used the final schema to code all the papers. The final coding scheme is shown in the results section.

Collecting Authors' Keywords from Studies

Furthermore, by examining the frequency and distribution of authors' keywords across the published papers, we can gain insights into the most common topics and themes that have been explored in the area of temporal educational research. We acknowledge that relying on keywords does not accurately represent all dimensions of the published research (e.g., method and conclusions); however, we feel it shows the main characteristics of the research in the area of temporal educational research from the authors' perspectives. It is worth noting that the trend of illustrating authors' keywords is commonly seen in mapping studies, which aim to provide an overview of a particular field or research area (Mohabbati et al., 2013; Petersen et al., 2008). Overall, while keywords are not a perfect representation of the research, they can still provide valuable information about the overall trends and characteristics of the field.

4.2. Results

The 176 included sources were published between January 1, 2017 and December 31, 2021.We structured the results section as follows. First, we address RQ1, this study organized four sections that separately discuss the components of RQ1. Next, we addressed RQ2 by providing relational visualizations for the pairs of RQ2.

4.2.1. Identified Research Questions Codes and Their Distributions

The result from the qualitative thematic coding shows 7 codes for the focus of studies' research questions (Table 4-5). Starting with *exploring socio-dynamic* which captures the dynamic of interaction patterns among peers during the discourse. The next code aims to *develop a method* or improve the existing ones. This code also includes proposing a methodological framework. The next code can be also considered as a subcategory of *method development* where the studies specifically aim to identify students *at risk* of failure. The next code directs the research question to *group* the users based on their behavior or performance. Two more codes, including *exploring SRL processes* and *identifying non-SRL learning indicators*, rely on the theoretical exploration of learning phenomenon.

Table 4-5. Focus of the research questions being asked in the papers.

Research Questions Focus (label)	**Description**
Exploring socio-dynamics	Analyzing the peer interactions and social dynamics during asynchronous discussion or collaborative tasks.
Method or algorithm development	Proposing or improving existing algorithms, methods, or frameworks. Also, authors can provide a novel framework that includes data collection, cleaning, and analysis approach. Furthermore, the study can compare the affordance of different analytical techniques.
At-risk student identification	Predicting students at risk of failure (drop out) by using a set of features and prediction model (The code is a subcategory of method development).
Group emergence/ group comparison by performance	Categorizing the users based on their online behavior or comparing the group of poor performing students vs high performing ones.
Exploring SRL processes	Identifying and exploring SRL-associated behaviors or engagement with materials.
Non-SRL learning indicators identification	Finding the indicators that can represent learning phenomenon which needs to be backed by learning theories (excluding SRL theory).
Time to intervention	Identifying the proper time for feedback or intervention

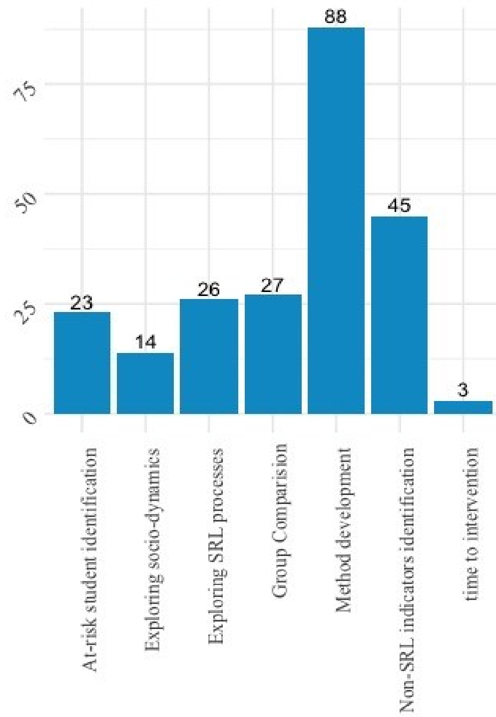

Figure 4-3. The distribution of asked research questions

Figure 4-3 presents the frequency of the research question codes in our corpus and their occurrence over the five-year period. Overall, 226 codes were assigned to 176 papers. The highest RQ focus was *method development* or *proposing methodological framework* (n=88). This suggests that the mainstream trend in educational temporal studies in the period of 2017-2021 was methodological development. The next trend in RQ focus in temporal studies is *exploring behaviors*, which can be an indicator of learning but are not based on SRL theory (n=45). In this category, studies relied on other theoretical background and learning constructs to justify discovered learning phenomenon. Aiming to group users based on their online behavior or performance (n=27), *exploring SRL-associated behaviors* (n=26), and *identifying students at the risk of failure* (n=23) are the next frequent categories, respectively. The least trending focus is to identify when it is the time to intervene to provide constructive feedback (n=3).

4.2.2. The Types of Data that Have Been Obtained and Engineered

Due to the differences in granularity levels that may appear in identified codes for data types through thematic analysis, we incorporated Knowledge Engineering (KE) to define three levels of codes. Knowledge engineering structures the sophisticated nature of data with different levels of granularity. The method has been mainly used in the AI field to model abstract observations to feed into a machine for a solution (Studer et al., 1998). The KE method focuses on structuring knowledge by capturing different aspects of it and can contribute to distinguishing different types of knowledge required for solving a specific task (Akkermans et al., 1994). Utilizing Knowledge Engineering, we systematically analyzed and categorized the data, focusing on the complexity and processing stages it undergoes. This process involved identifying the inherent characteristics of the data at various stages, from its initial collection to the more abstracted forms. This approach inspired us to structure different data types used in temporal techniques (Table 4-6). Based on hierarchy composition, we proposed three levels of data including: 1) Raw level data, where the data type is collected and stored at the lowest level; 2) Feature level, which refers to features extracted from the raw data; 3) Pattern level, representing the engineered pattern from the features to represent the temporality aspect. Using these levels, we were able to distinguish different types of utilized data in our corpus.

Table 4-6. Three levels of the utilized data.

Data types	Description
Raw level data	The data at the raw level of collection
Feature level	Processed raw data into features/variables
Pattern level	Engineered or extracted temporal patterns from the features

Table 7 describes the identified raw data, and Figure 4-4 shows the distribution of identified raw data in our corpus. As can be seen, the most utilized raw data is LMS log data (n=103), which records the activities of the users. Performance measures, consisting of information associated with the performance outcome of the user (e.g., course grades and quiz marks), is the second most incorporated data (n=52). This type of data is essential for studies focusing on the performance prediction of students (e.g., Q. Hu & Rangwala, 2019; Morsy & Karypis, 2019; Van Goidsenhoven et al., 2020). Next are learning products (n=38) and customized software log data (n=35). The rest of the raw data types were substantially less frequent than the aforementioned raw data.

Table 4-7. Identified raw level of obtained raw data

Raw data	Description
LMS log data	All recorded activities from students' interaction with LMS.
Customized software log data	Specific software designed for interaction (e.g., adaptive learning system integrated with an intelligent tutoring system).
Learning product	Instances that represent products of learning such as discussion content, programming code, or essay.
Performance measure	Assessments of learning (e.g., course grades, quiz marks).
Contextual data	Contextual or recorded context. Examples include interview context.
Self-reported	Self-assessed or self-reported survey data.
Learner characteristics	Students' administrative information such as background, prior courses, gender.
Multimodal	Different modalities such as video records, eye tracking, EEG, mobile, tablet, PC, location tracking, and others

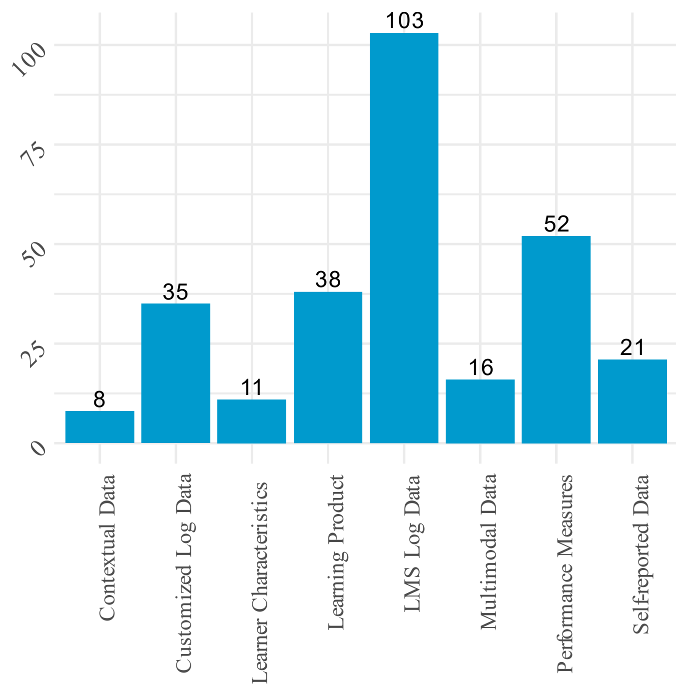

Figure 4-4. The distribution of raw data

The second data type is feature-level data, where raw data is often engineered into sets of features that can address specific research questions. Table 4-8 presents the discovered codes from the studies, while Figure 4-5 illustrates their distribution. At this level, three types of codes—event, time, and traces—are paramount. Traces represent “how a learner operates on particular information at a point in time and in a relatively well-identified context” (Winne, 2020). In other words, trace data include records of activities that have meaning based on a theory. A list of these traces is also provided in Table 4-8. The distribution of codes shows that event features are dominant (n=164), as studies overwhelmingly collect records of clickstream from users, except in cases where studies relied solely on survey and self-reported data. Next, time features were also vastly incorporated into temporal analysis (n=60). An example of this feature is when studies focus on analyzing intervals, timely behaviors, or incorporating time features in their prediction models (e.g. Chen, 2021; Lwande et al., 2021; Saqr & López-Pernas, 2021; Shin et al., 2021). Regarding traces, discussion forum activities have the highest trend (n=48), followed by reading (n=39). Other trace activities such as highlighting, viewing a dashboard, and note-taking were also utilized in a total of 50 studies.

Table 4-8. Feature level data and their descriptions.

Feature level	**Description**
Events	Any record in a log file that has a specific meaning
Time feature	Incorporating time as an independent feature in the method (e.g., the time taken for a student to answer, lag time, the time interval between adjacent learning activities, time series)
The traces:	according to Winne (2020), “traces represent how a learner operates on particular information at a point in time and in a relatively well-identified context” (e.g., highlighting task, goal setting, practices)
Trace-exercise	Trace data from users engaging with exercise activities
Trace-reading	Access to reading content or accessing content
Trace-quiz	Activities including quizzes and assignments, viewing quiz results, quiz pages
Trace-video	Watching learning-related videos
Trace-forum	Forum activities such as posting, replying, and reading others' posts

Trace-feedback	Asking the user to provide feedback after an activity
Trace-other	Any other trace activities such as highlighting tasks, accessing dashboards, setting goals, note-taking, playing games

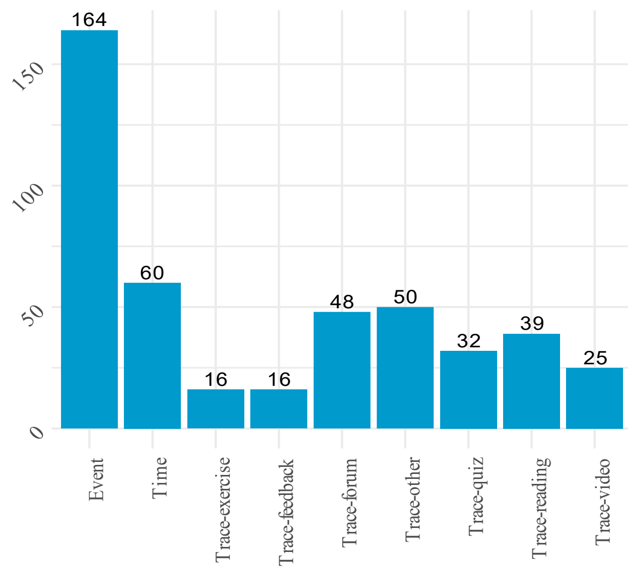

Figure 4-5. The distribution of feature level data.

The next level of data includes the pattern level, which is often engineered from feature data and can be honed through interplay with research questions and analysis. In other words, studies often generate pattern-level data to reveal patterns in their data, addressing their research questions and bringing insight into their findings. The list of pattern-level codes can be seen in Table 4-9, and their distribution is shown in Figure 4-6. The most frequently engineered pattern is the *summative pattern* (n=72), such as the sum of the number of users' logins during a week, which was mainly utilized for statistical tests. The next most prevalent pattern is the *transitional pattern* (n=67), where studies aim to understand transitions between states (e.g., online activities as states: watching video → reading content). *Other sequential patterns* not belonging to any specific pattern level constitute 33 in our corpus. Following these are *Event sequences*, which consist of blocks of sequential instances (n=28) followed by *group event pattern* (n=22). Lastly, 11 studies did not include an engineer pattern-level data in their research.

Table 4-9. The engineered pattern-level.

Pattern level data	Description
Event sequence	A sequence of events within a defined period that describes the phenomenon. In this sequence pattern, the structure, including the length and composition of the sequence, is the primary concern.
Group event pattern	Groups of sequences that share similarities in transitions or frequencies of specific events.
Transitional pattern	The transition from one event to the next is a primary concern. This pattern is commonly used for process mining techniques and lag analysis.
Other sequential patterns	Any event sequences not belonging to the three categories above, e.g., in network analysis, the path through network nodes, or a novel engineered sequence.
Summative features	Mainly averages, means, and frequencies.

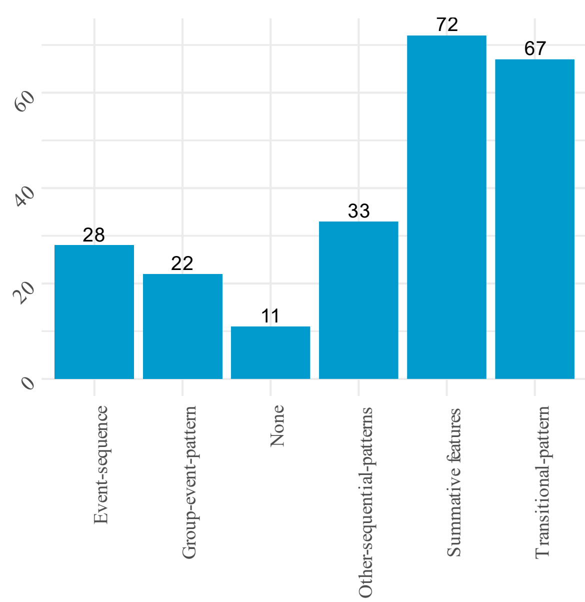

Figure 4-6. The distribution of pattern level data.

4.2.3. Utilized Analytical Techniques

This study identified 10 groups of analytical techniques used in temporal educational research. Table 4-10 describes the full set of identified codes; overall, 300 codes were assigned to 176 papers (as a paper can receive multiple codes). Figure 4-7 shows the frequency of the codes for the analytical techniques and their distribution over the five-year period. The descriptive analysis of the codes indicates that *process mining* (n=70) and *visual analysis* (n=62) are the most frequently utilized techniques, followed by *statistical analysis (stat)* (n=43), and *cluster analysis* (n=39). The high trend in the use of process mining suggests the affordance of this technique to reveal temporal dynamics of the behaviors. Studies often interpreted the identified temporal behaviors as a study strategy or learning engagement pattern that is explained by learning theories. Interestingly, the high use of *visual analysis* in temporal studies can show the importance of visualization to discover and explain the dynamicity of behaviors.

Table 4-10. Identified codes for analytical technique.

Analytical Techniques	**Description**
Process mining	Process mining detects the significance of the transitions between events. Some examples are lag analysis, fuzzy miner, inductive miner, etc.
Frequent sequence mining	Different from process mining, this technique detects frequent sequences of events that occur more often during defined period. For instance, this technique also looks into the whole sequence of activities during a week and compares the weeks.
Cluster analysis	Clustering techniques group data points based on statistical similarity and are usually followed by statistical analysis to identify the differences between clusters.
Text mining/Content analysis	Text mining or Content analysis is defined as the use of any natural language processing technique to model contextual data.
Neural network	The technique is the use of the network of neurons to implement a prediction model. Any type of deep neural network is considered in this category.
Qualitative analysis	Qualitative techniques are used to qualitatively examine and/or discuss the nature of the phenomenon.
Basic statistical analysis	Any statistical standalone test that is not part of another technique (e.g., comparing clusters). Examples include correlational test, ANOVA, pre-post test, entropy analysis, interaction over time, time window analysis.

Network analysis	The aim is to identify and structure the relations to explain social phenomena using nodes and relation lines.
Visualization analysis	The main aim of visual analysis is to communicate the meaning of data through visualizing data. We focused on the explanatory power of visualization as this code is assigned if the use of visualization is crucial to driving conclusions in the research. An example is that the researcher uses visualization to compare two phenomena to identify a pattern(s) and drive a conclusion.
Other prediction models	This includes any other techniques used to develop a prediction model (e.g., random forest, SVM)

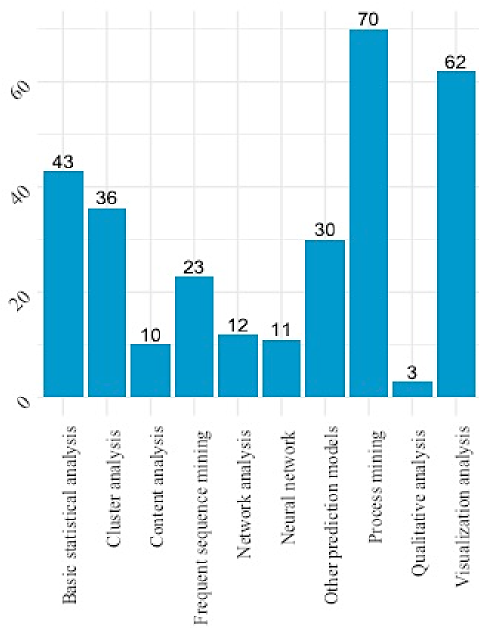

Figure 4-7. The distribution and trend of the utilized techniques.

From a temporality perspective, some analytical techniques work exclusively with time data (process mining, frequent sequence analysis) while other techniques are more general. In temporal educational studies, the more general techniques, such as statistics or clustering were either applied to the outputs of the process mining or frequent sequence mining, or they were applied to features capturing temporal aspects of data, e.g., frequency of learners' actions within a time window. Often studies utilized several techniques together. We presented this cross relation between them in Figure 4-8. where the main diagonal shows the number of times the technique was the only technique used in the study, and other cells show counts when techniques were used together. We showed that the high trend in using visualization analysis indicates the crucial role of this technique to reveal temporal aspects, and it seems that, without extensive visualizations, studies would not be able to derive their findings. Therefore, this technique was extensively utilized with other techniques, especially process mining. The second most utilized technique was process mining, one of the ‘pure’ temporal techniques used in the studies. As Figure 4-8 shows, when process mining was used, it was used by itself only

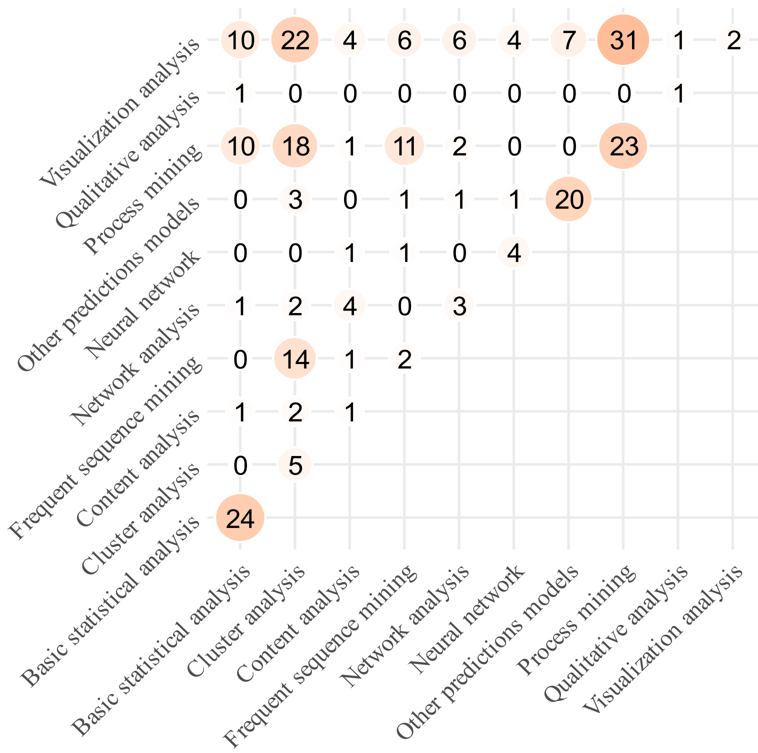

Figure 4-8. Analytical techniques being used together. The main diagonal shows the number of studies where the technique was the only technique used.

in 23 studies. More often it was used with other techniques, such as with visualization to interpret the process models (n=31), cluster analysis (n=18) to either cluster students based on some characteristics or to cluster discovered processes, frequent sequence mining (n=11) was often used in parallel with the process mining or processes of frequent sequences were investigated, and basic statistical test were used to investigate other aspects of the students learning behaviors (n=10).

4.2.4. Insight About Learning

Table 4-11 shows the identified codes for insight about learning. Overall, 212 codes were assigned to 176 papers. Figure 4-9 shows the frequency of the codes and their distribution over the five-year period. The highest learning insight trend is *identifying indicators of learning* (n=77). The next highest occurrence in this section refers to the *no-learning-focus-outcome* (n=51) that the studies did not (sufficiently) show the circumstances of the learning phenomenon. These studies often focused on developing a method rather than examining the impact on learning. From the time progression chart (Figure 4-9, right) we can discern a drop from a high in 2017 in papers contributing insight on collaboration, and a spike in 2019 for studies with no learning focus.

In the next section, we will further discuss the association between the focus of RQs, the utilized techniques, and learning insights. Overall, our identified learning insights suggest that three codes are user-centric including *Learning Indicators*, *Collaboration*, and *Time-and-Learning*. These codes reflect how the behaviors of students impact their learning. Two other codes, *Course-design* and *Feedback*, are instructor-centric codes that imply the role of the instructor to intervene or design learning materials to impact students' learning.

Table 4-11. Identified codes for insight about learning.

Insights about learning	**Description**
Course design	The researcher shows specific course design can impact learning. This also includes scaffolded design experiments.
Learning indicators	The researcher identifies a set of theoretically grounded indicators that can characterize learning.
Feedback	The study finds the effect of feedback on learning.
Collaboration	The study discovers the effect of collaboration on knowledge building. This also includes investigating the progression of an idea, the quality of the idea, or the statistics of interactions during discussion. Studies often investigate how the group of users collaborate to reach a goal.
Time on learning	The researcher shows and discusses the effect of time on learning.
No learning focus outcome	The study does not provide sufficient justification for showing how learning happens or any impact on learning.

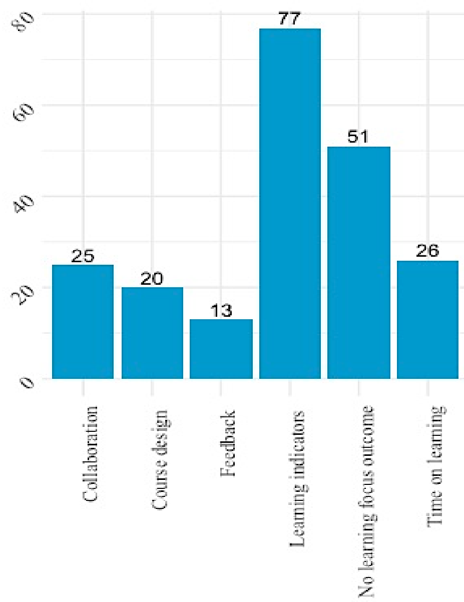

Figure 4-9. The distribution of insights about learning.

4.2.5. Identifying the Associations Between the Research Questions Being Asked, the Analytical Techniques, and the Insights About Learning

In this section, first, we explore the associations between the focus of the research questions and the utilized analytical techniques (Figure 4-10), and then further details will be discussed by adding the dimension of learning insight (Figure 4-11). Figure 4-10 shows the relationship between the research questions being asked in papers crosslinked with the techniques utilized to address them. The x-axis shows our identified codes regarding research questions being asked, and y-axis represents the codes regarding techniques, and each circle shows the number of papers that map to a particular RQ that are addressed with a particular technique.

As already discussed (section RQ1-(a)), aiming to develop a method is the most common research question focus. In this category, utilizing *visualization technique* (n=35), *process mining* (n=29), *other prediction models* (n=23), and *clustering* (n=21) are the most trending techniques, accordingly. The figure also shows that *process mining* is a viable technique for all types of research questions except for identifying students at risk of failure. The high trend in utilizing process mining suggests that the process model, the output of the process mining, can characterize temporal patterns. This means that any behavior changes can be measured and interpreted based on underlying theory. In other words, the theory defined the meaning of each state of a particular behavior (e.g., clicking on video content, posting a discussion), and *process mining* measures the transitions between states (e.g., from viewing a discussion -> to watching a video). Studies often visualized and interpreted the transitions to infer how learning happened. Moreover, some studies incorporated *clustering technique* besides *process mining* to provide a deeper comparison between behaviors (Boroujeni & Dillenbourg, 2019; Fan & Saint, 2021a; Huang & Lajoie, 2021a). Similarly, *frequent sequence mining* generates sequences, composed from different actions or states, with frequencies in the defined period. Therefore, the technique provides strong explanatory power, especially to show how the users interact with the learning management system to reveal SRL and non-SRL associated activities. For instance, a study conducted by (Jovanović et al., 2017a)) utilized this analytical technique to unveil the temporal behavior that can be associated with the SRL phase in flipped classroom settings. Furthermore, it is posited that frequent sequence mining and process mining can complement each other (Chen et al., 2017c), and a study showed how these techniques can reveal different aspects of temporality in SRL associated behaviors (Matcha, Gašević, Ahmad Uzir, et al., 2019). On the other hand, to identify *at-risk* students, the main focus is to achieve a high accuracy prediction rate through incorporating temporal features. Therefore, this category chiefly employed prediction models (n=18), consisting of *neural network* (n=3) and *other prediction models* (n=15), to address their research questions.

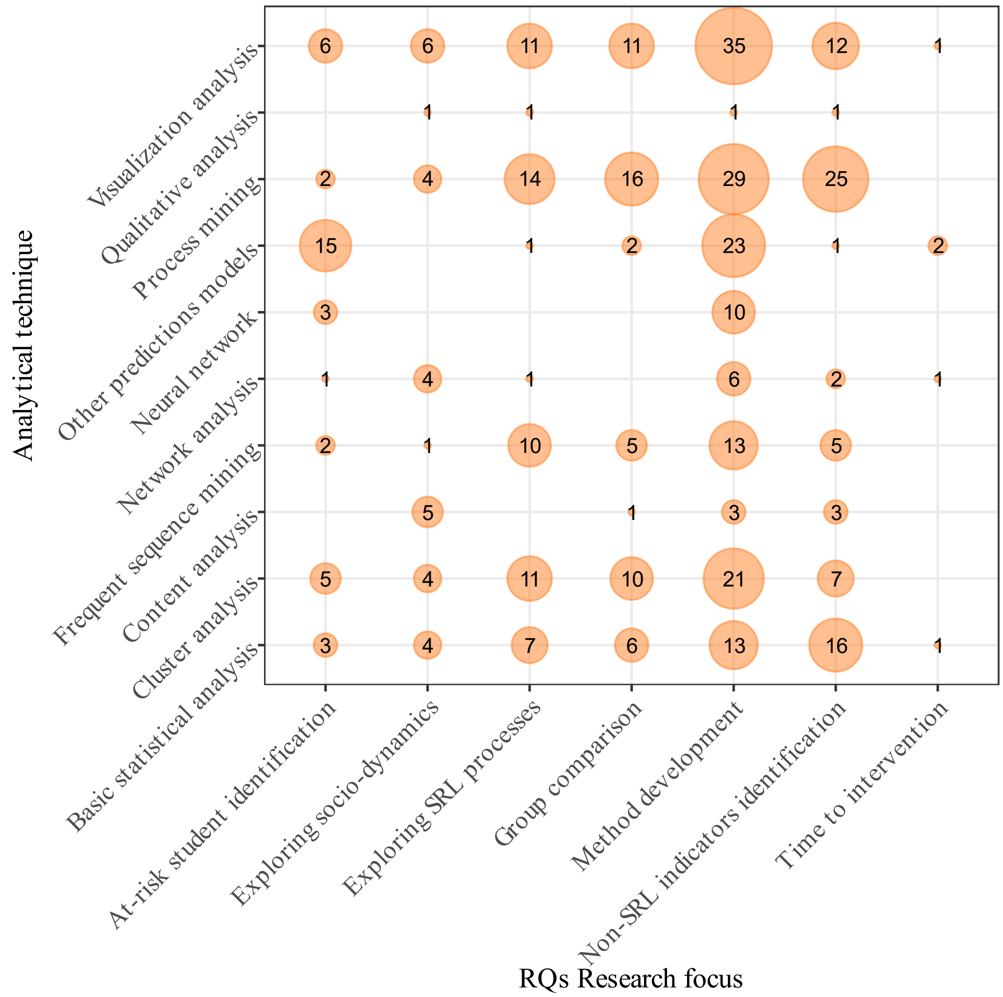

Figure 4-10. Relationship between asked research questions and utilized techniques

Further analysis by considering the codes for learning insights (Figure 4-11) reveals the trend in the association of RQs' foci and analytical techniques based on inferred insight about learning. The plot is divided based on the revealed insights about learning, and the x-axis represents our identified codes regarding research questions being asked, and the y-axis shows the codes regarding techniques, and each circle shows the number of papers that map to a particular RQ that are addressed with a particular technique respecting revealed learning insight.

Starting with capturing *indicators of learning* (user-centric insight), where constitutes the highest attentions of research foci, studies with the focus *of developing a method*, mainly utilized *visualization* (n=15), *process mining* (n=14), *clustering* (n=9), and *frequent sequence techniques* (n=8). Studies in this category often developed a methodological framework to generate sequences of activities based on underpinning theory to reveal the dynamicity of learning phenomenon. In this learning insight, the main difference between *exploring SRL processes* and *exploring non-SRL learning indicators* was that SRL studies substantially used more *frequent sequence mining* and *clustering techniques* (n=8, n=9), in comparison with non-SRL studies (n=3, n=4). The comparison suggests that tools such as TraMineR (Gabadinho et al., 2011) that are based on *frequent sequence mining* techniques are popular to create sequences of activities that are associated with SRL processes. Then, these activities can be clustered to characterize and compare the students' behavior. *Content analysis technique* is not used frequently; it was used most often (n=3) for *exploring non-SRL learning indicators*. Finally, studies that were concerned with identifying students at risk of failure and identifying the time to enter intervention are more action-oriented and did not result in revealing learning indicators.

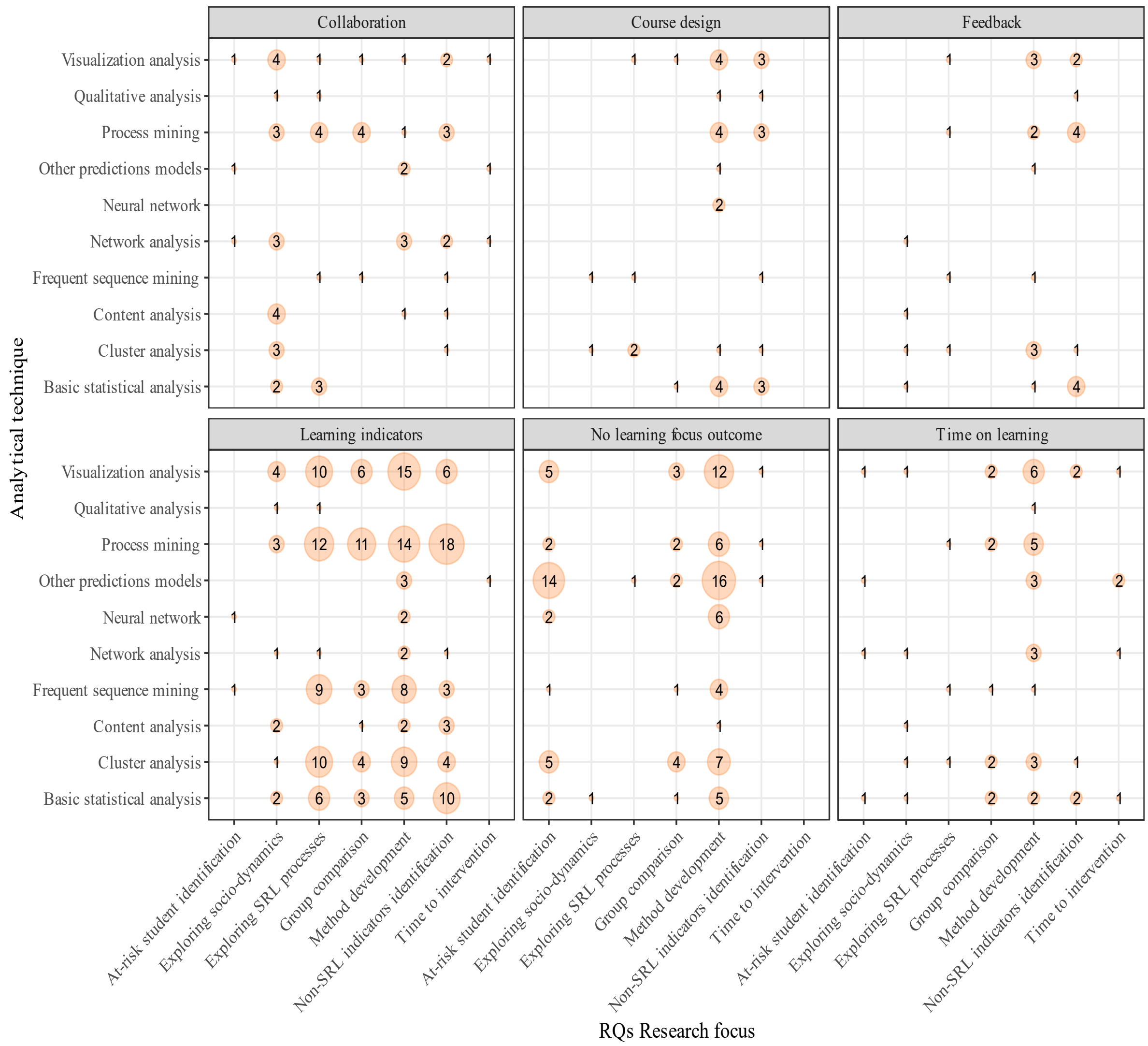

Figure 4-11. Relationship between research question foci and analytical technique respecting learning insight.

Two other user-centric insights, consisting of *collaboration* and *time on learning*, have a distinctive trend in terms of the foci of RQs and the utilized techniques. Studies that illustrate the impact of collaboration in learning, focused on *exploring socio-dynamic* and mainly utilized *text mining* (n=4), *visualization* (n=4), *process mining* (n=3), and *network analysis* (n=3). The instance of these types of studies is tracing the progression of the idea through online discourse (S. Liu et al., 2021; M. Wang et al., 2020). *Network analysis* was utilized relatively more in *collaboration*. It is likely that the authors reported using this method to show the connections of interactions through discourse. This allowed them to follow how adding a new idea can trigger higher discussion activity (Lee & Tan, 2017; N. Sher et al., 2020). Overall, the technique can provide a deeper understanding of the construction of collaboration. On the other hand, studies that inferred the impact of *time on the learning* process had *method development* as RQ focus through mostly using visualization (n=6) and process mining (n=5).

Two instructor-centric insights (*feedback* and *course design*) demonstrated a similar pattern, that *method development* and *exploring non-SRL indicators* were the highest RQs foci. In *course design*, authors often proposed a new framework for learning and conducted experiment to explore the impact of their proposed method on users' behavior, mainly using *process mining* or *basic statistical tests*. Feedback also follows a similar rationale to examine the impact of feedback.

Lastly, studies *without learning insight focus* outcomes mainly focused on *developing method* and identifying *students at risk* of failure. These types of studies extensively focused on methodological description to improve or create a novel approach to address their research questions. These types of studies are often found in the area of educational data mining (EDM), which is more algorithm-centric and less attention is given to studying impact on learning. In our corpus, EDM constituted 15 papers that 9 of which were coded as having no learning outcome focus. Overall, papers without learning insight aimed to improve the performance of the existing model by utilizing a new set of temporal features or proposing a new algorithm based on temporal data (n=46 out of 51). Notably, deep neural networks are gaining attention in this category.

4.2.6. The relationships between data being collected and analytical techniques and discovered insight about learning

Similar to the prior approach, we provided three-dimensional plots that each show the associations between the data obtained, the techniques utilized, and the insights about learning that were discovered. Each plot is categorized based on a particular insight, where the x-axis represents the data type, the y-axis shows the utilized analytical techniques, and each circle represents the number of papers counted based on data, technique, and insight. Since we have three levels of data, we generated three different plots.

Starting with the first level of data, raw data, it can be seen (Figure 4-12) that the use of *process mining* and *visualization* with LMS data were the predominant patterns across all learning insights. In user-centric insights, the next trending data (after LMS data) are *customized log data*, *learning products*, and *performance measures*. Learning indicator insights incorporated mostly these data as well as *self-reported data*. On the other hand, studies mainly relied on *learning product* data (besides *LMS data*) for gaining *collaboration insights*. The insight of *time on Learning* relied on performance measures data alongside LMS data.

Two instructor-centric insights including *course design* and *feedback*, which pertain to the instructor's role in intervening or designing learning materials to influence student learning, exhibit a similar overall trend. However, *course design* insight had more studies using *customized software log data*. In this category, basic *statistical* tests (n=3), *process mining* (n=2), and *visualization* (n=2) were the most utilized techniques.

Studies without a learning focus outcome mainly collected *LMS data* and p*erformance measure*, and they predominantly utilized prediction models (including neural networks and other prediction models).

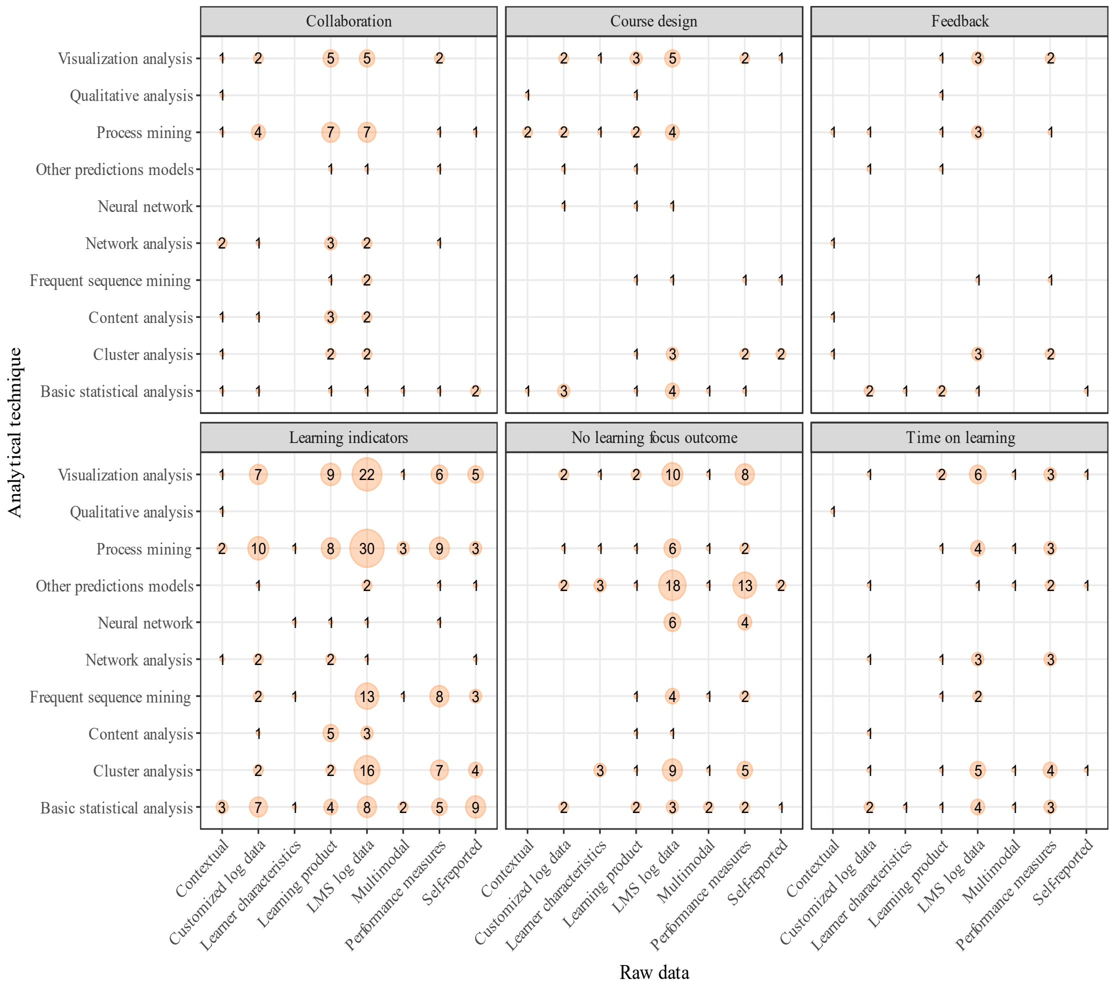

Figure 4-12. The associations between raw level data, techniques, and insights.

Exploring the next relational (Figure 4-13) indicated that *event* was the predominant feature level in studies. Besides *events*, the use of *trace data* is apparent in studies with learning insight. In the *learning indicator* which is the highest inferred user-cantic insight, as already discussed in our RQ1, *process mining* was the most trending technique and using *trace-other* (n=21), *trace-forum* (n=19), *trace-reading* (n=17) were the highest trending feature level. In collaboration, *trace-forum,* and *trace-other* the domination after *event* data. In this category, studies often use collaborative knowledge building theory to observe how users' communication between users led to learning as well as to explore how dynamic collaboration and knowledge building happens (e.g. (Engerer, 2020; Sobocinski et al., 2017)). To manifest these communications, the most utilized techniques were *process mining*, *network analysis*, and *visualization*. In the next user-centric insight, *time on learning*, the high use of *time* as the feature (after *event*) is noticeable. In our corpus, we often coded *time* and *event* alongside, because studies used *event* to store time-related information such as when the user logged in to the system during the week. For an instance, a study aimed to gain insight into how users dedicate their attentions to information processing using a time-driven approach (counting the number of particular attentions with respect to time-lapses), and then assess their attention with their performance (Poitras et al., 2021).

In *feedback* and *course design* insight, different *trace data* are roughly

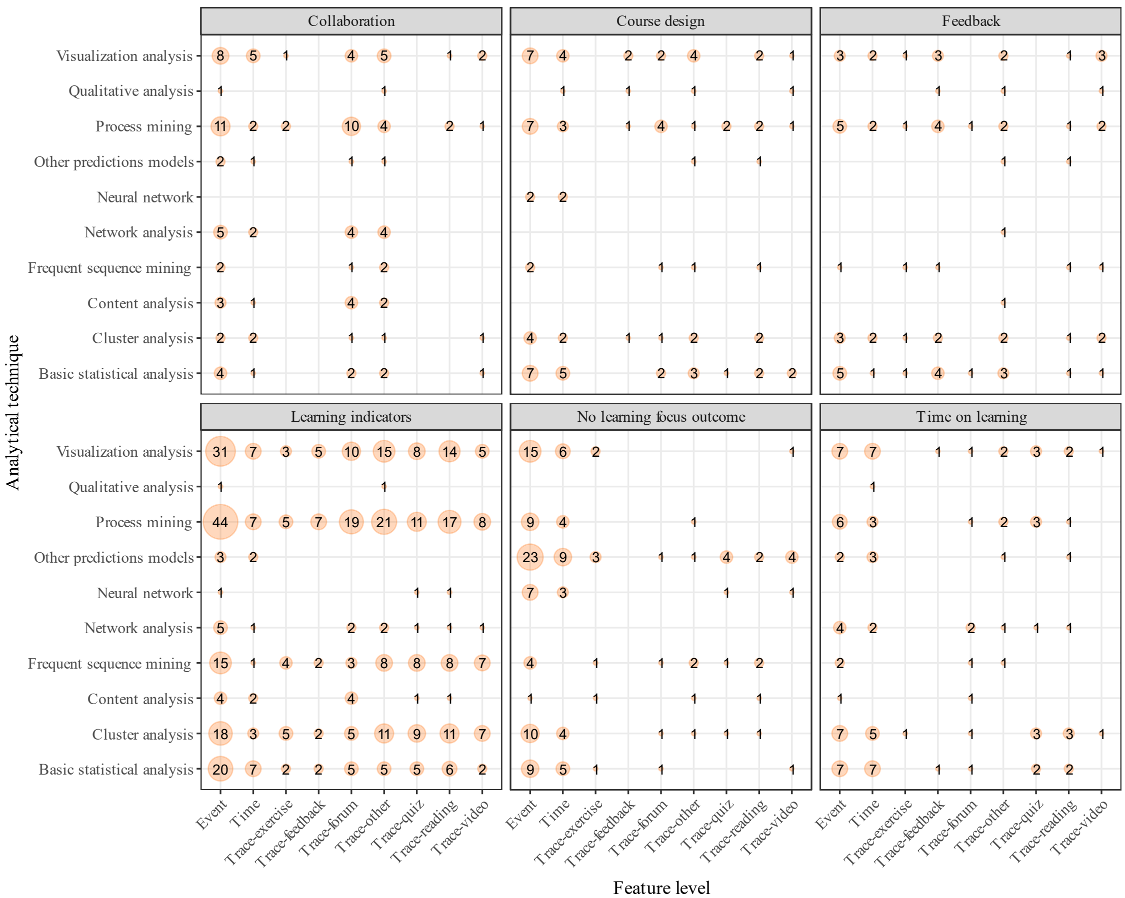

Figure 4-13. The associations between feature level data, techniques, and insights.

distributed; besides, it is noticeable that studies used *clustering* and *statistical* test besides *process mining* and *visualization*. An example of gaining these insights is that the study observed how certain instructor feedback can impact students' performance in educational gameplay (Yang & Lu, 2021), and regarding the impact of *course design*, the study explored the temporal aspect of learning by comparing students who received scaffolded instructions versus non-scaffolded instructions, using process mining (Lämsä et al., 2019).

On the other hand, trace data were used less in studies *without learning focus outcome*. The main reason is that studies in this category allocated less or no attention to theory, rather they focused on the method or algorithm development.

The last relational illustration is shown in Figure 4-14 which considers pattern level data with technique and insight about learning. As it can be seen, the strong association between *process mining* and the *transitional pattern* is apparent in all insights. Specifically, studies that gained insight about *learning indication* and *collaboration* mostly engineered *transitional patterns* to use for *process mining* and *visualization*. Furthermore, *text mining* using *summative* pattern is comparatively higher in *Collaboration* than in other insights.

In contrast, *other sequential pattern* was utilized less in studies with learning insight but mostly in studies without learning insight. In this respect, studies often proposed a method to engineer a set of novel patterns level for prediction algorithm to improve their performance (e.g. Fatahi et al., 2018; Q. Hu & Rangwala, 2019; Mahzoon, Maher, Eltayeby, Dou, et al., 2018; Qiao & Hu, 2020; Wu et al., 2019). Similarly, in the area of knowledge tracing in this category, studies proposed novel methods quipped with the engineered pattern for knowledge tracing without or less attention to elaborate on the impact of their method on learning (e.g. (Choi et al., 2020; Pandey & Srivastava, 2020)).

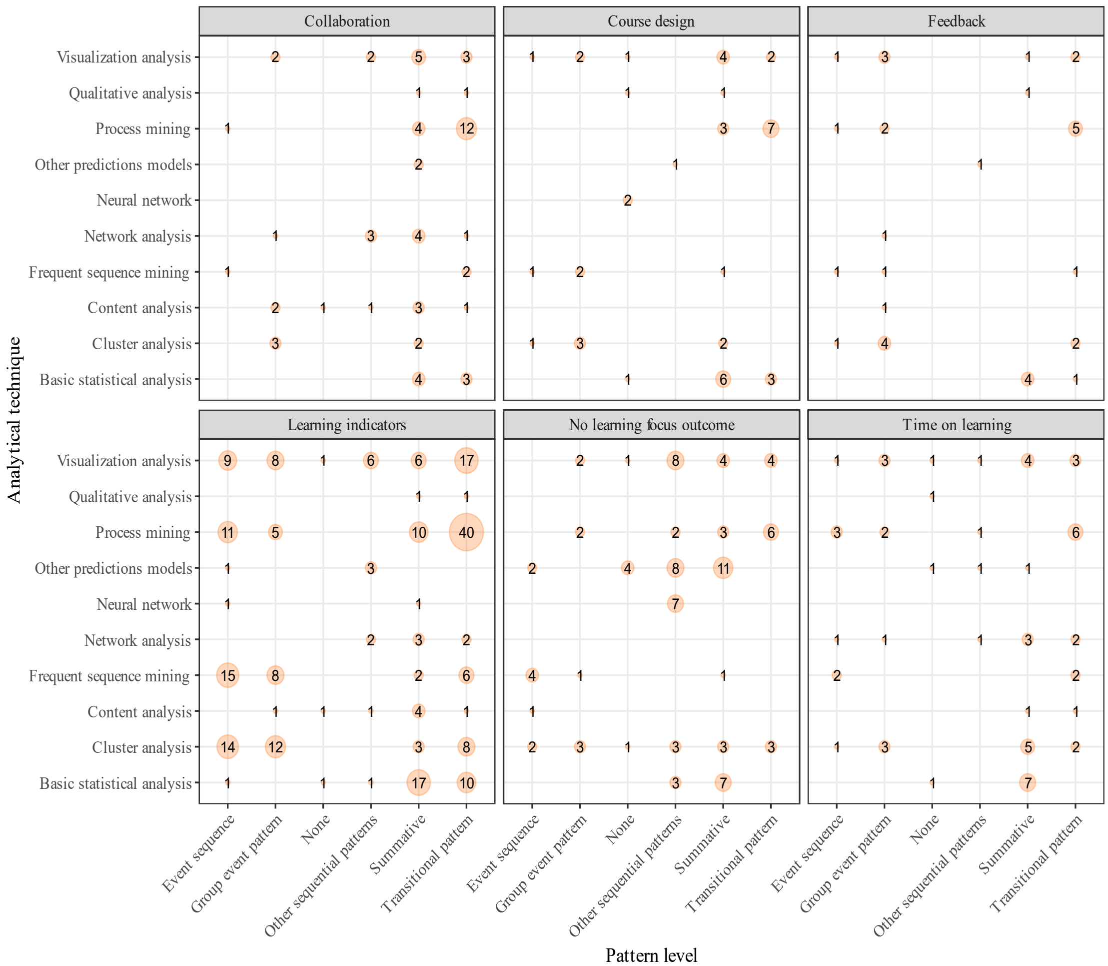

Figure 4-14. The associations between pattern level data, technique, and insight.

4.3. Discussion

Learning is a process that happens over time. The circumstances of the learning process can provide insight into the understanding of the learning phenomenon. Temporal analysis is the field dedicated to exploring the learning process and its temporality. In recent years, the temporal aspect of learning has received increased attention in the learning analytics community, and studies utilized several methodological techniques to exploit temporal information. However, despite research efforts to date, it is not clear what the associations are between asked research questions, utilized techniques, and inferred insights about learning. Therefore, in this study, we investigated the affordance of temporal techniques and showed how authors used them to reveal learning.

The findings in this mapping study can help orient and guide researchers in preparation for conducting their temporal studies by providing a list of relevant works that can lead them to selecting proper techniques based on their research questions and what type of insight they are anticipating. For this purpose, before conducting a study, researchers can start their investigations by exploring the lists of published temporal studies in different categories (provided in Appendixes 1 and 2). Starting with the type of research questions asked, researchers can look up research question types from what we provided to see which are closely related to their inquiries. To illustrate, researchers interested in investigating learning indicators for SRL processes using temporal approaches can quickly identify and examine the list of 22 studies for closer examination, gaining an overview of the state-of-the-art, and helping them to select appropriate techniques and data features to answer their research questions. For example, they can choose a set of papers that developed a sequential model to characterize learning strategies (Fan & Saint, 2021; Jovanović, Dawson, Joksimović, & Siemens, 2020; Jovanović et al., 2017; Saint, Fan, Singh, Gasevic, & Pardo, 2021). These papers defined a learning strategy as *“Any thoughts, behaviors, beliefs or emotions that facilitate the acquisition, understanding, or later transfer of new knowledge and skills”* (Jovanović et al., 2017a). Learning strategies define how students use a different sequence of activities that show the characteristics of an individual’s learning. They can then compare the approach in these papers with approaches used in another study, where researchers utilized various techniques to explore the temporality of learning strategy and compared how the results from each technique agreed (Matcha, Gašević, Ahmad Uzir, et al., 2019).

Secondly, we provided a list of inferred insights about learning that can help researchers to explore their anticipated insight. Appendix 2 helps researchers to locate studies that focused on particular learning insight from the research question perspective, and what techniques were used to accomplish it. As we discussed earlier, the most prevalent learning insight from temporal studies was to identify learning indicators in order to develop a method to characterize online behavior of users. In this category, the studies often defined a set of activities that are associated with the theoretical background, and then identify temporal changes in the activities or interpret the sequences of activities as learning progression. For instance, studies identified a certain sequence of students' activities to be associated with an SRL phase (e.g., enactment of learning tactics), and the recurrences of the phases to indicate learning progression (Fan & Saint, 2021a; L. Huang & Lajoie, 2021a; Jovanović et al., 2020; Wang et al., 2021). After learning indicator insight, the second largest group of temporal studies were not aimed toward theoretical insights from the perspective of learning theories. These types of temporal studies often harnessed the predictive power of temporal features (e.g., time and order of activity) for their proposed model; contributing new algorithms or proposing a set of new (temporal) features to improve the performance of their model.

This study also showed that obtained data can have three granular levels. At the *raw data* level, the highest attention was given to LMS log data. This *raw data* is typically processed and transformed into events and trace data at the *feature level*. Our findings are aligned with studies that highlighted the importance of events in educational studies as they asserted that, the trend in data, has shifted from aptitudes, which mainly relies on interviews and self-reported surveys, to event-based studies that trace activities of students to derive learning occurrence (Ding et al., 2019; Panadero, 2017; Winne, 2014; Zhou et al., 2010; Zhou & Winne, 2012).

The next level of the data, i.e., the *pattern level*, requires a data engineering process to model the temporal behavioral pattern of the users. The outcome feeds into the temporal technique for analysis. Our finding showed that the mainstream *pattern level* data is the *transitional pattern* that was utilized for the *process mining* technique. This technique was dominant in our corpus that shows the tendency of the body of temporal studies to reveal temporal behavior. The reason for this tendency is due to *process mining* affordance to reveal temporal behavior that has been extensively elaborated in the literature. As an instance, Wang and colleagues proposed a framework that utilized process mining to interpret and evaluate the temporal behavior of students in assignments. They also provided a case study that aimed to model students' assignment activities to show the process of students' study effect (Y. Y. Wang et al., 2019). Similarly, other temporal studies emphasized the affordance of process mining to infer indicators of learning (Bannert et al., 2014; Jo et al., 2014; Neyem et al., 2017; Saint et al., 2021; Sun et al., 2021; Umer et al., 2019; Uzir et al., 2020).

Furthermore, our findings showed that when conducting temporal studies researchers often use a combination of techniques. Some techniques work exclusively with the time data, namely process mining and frequent sequence analysis. These two techniques differ in several important ways and are complementary in what they can uncover (Chen et al., 2017c). Frequent sequence mining finds concrete sequences of learning actions that can be directly observed in individual students' log files or higher-level derived constructs, such as SRL phases. As a result, the presence of these sequences can indicate a student belongs to a particular group or demonstrates certain characteristics, potentially leading to intervention. The models that are outcomes of the process mining techniques are probabilistic in nature, specifying frequencies or probabilities of transitions between steps in the learning process, such as frequency of transitions between course activities. Although such models allow us to understand the underlying learning process, they are generally unsuitable for relating individual student activity to the discovered models. Visualization techniques, through their affordances, have the power to show temporality by depicting steps of learning activities as they unfold over time. However, the visualizations were used in this capacity quite rarely. They were often used in combination with other techniques, as we detailed in the results section.

Other techniques used by temporal studies are more general and examine the temporal nature of learning using data features designed to capture temporality. For example, a study by Du and colleagues investigated the temporal pattern in engaging with learning materials by computing the time and physical location of the students and used statistical analysis to show the correlation with academic performance (Du et al., 2019). Another study used activity session feature, which included a trace log based on the 30-minutes threshold, and a clustering technique to differentiate groups of students with different level SRL behaviors (de Barba et al., 2020). As a potential direction for further analysis, our findings can be used as a source of literature to identify data features that capture temporality to examine particular research questions and learning insights.

4.4. Limitations

Our mapping study was associated with the following limitations. First, the papers were collected through a search on databases, and some journal websites might have less accurate search mechanisms. Furthermore, some of them did not support the search query in Table 4-3 (e.g., using AND, OR, and asterisk (*) operations). To address this issue, we manually inserted combinations of search terms (inserting the individual term one at the moment). Second, the relational analysis had redundancy and overlapping issues, which means that a paper can have several codes at the same time, and relational codes faced multiplication in numbers. This is the reason that relational numbers are more than distribution numbers. However, this issue did not deter showing the trend in associations between the asked research question, utilized techniques, and obtained insight. We also provided a cross-relational table to show the techniques being used together (Figure 4-8). Another limitation of the study is the covered period of five years, for reasons listed in section 2.1.1. Albeit we believe the codes provided in this study to be stable, we cannot claim this mapping study to be exhaustive, rather exploratory in nature. New codes may be uncovered if studies outside of the mapped period are included. Similarly, the uncovered relationships between research foci, analytical techniques, and learning insights are representative only of the period covered.

4.5. Conclusion

In conclusion, by providing a list of insights gained about learning, we demonstrated how temporal studies have been able to unveil learning processes using different analytical techniques. To address RQ1, we illustrated that research primarily concentrated on methodological development, followed by exploring non-SRL and SRL behavior identification. Data was categorized into three types: raw, feature, and pattern level. Raw data, with LMS log data being most common, was primarily used. Feature-level data processed these raw data into specific variables, predominantly recording event features. Finally, pattern-level data was utilized to derive insights from features, with summative patterns, such as the total number of user logins during a week, being the most prevalent. In terms of analytical techniques, the most popular were process mining, which tracks the sequence of actions or events, and visual analysis. Regarding learning insights, the top finding was identifying indicators of a person's learning progress. However, some studies focused not primarily on learning but on refining the methods themselves.

We also addressed RQ2, which focused on illustrating the associations between the components of RQ1. We showed that the primary focus of many studies is method development, and techniques like visualization, process mining, and various prediction models have been widely used to address this. Interestingly, process mining is used across many research questions and helps capture patterns of behavior over time. Some studies even combine multiple techniques, such as clustering with process mining, to draw richer conclusions. When it comes to understanding self-regulated learning, sequence mining and clustering seem particularly popular. However, if the goal is to identify students at risk of underperforming, predictive models take center stage. Looking at the relationship between data types, techniques, and learning insights, it's evident that process mining combined with Learning Management System (LMS) data is a dominant trend. Studies focused on collaboration often leverage text mining and network analysis. Those without a direct learning insight tend to emphasize method development, and deep neural networks are gaining traction in this area. Across these studies, the types of data, from raw to pattern level, and their associated techniques play a crucial role in shaping the insights derived about learning.

Overall, this chapter contributed to widening the understanding of current trends in temporal educational studies. We showcased the connections between the research questions posed by researchers, the data obtained, and the analytical techniques utilized, while considering the learning insights. This evolution of the field adds an extra layer to previous systematic or overview studies that addressed temporality aspects in educational research (Gašević et al., 2017; Knight, Wise, et al., 2017; Reimann et al., 2014) by providing the current trend in components of temporal studies. Knowing what techniques have been used can help researchers in two ways. First, it allows them to quickly identify effective techniques used previously, based on the similarity of their research focus, data in hand, and desired outcomes with past research. Second, it supports exploratory research by selecting novel techniques that were rarely utilized before, aiming to unravel different aspects of temporality. Furthermore, this study found that to provide learning insights, it is important to utilize techniques that are interpretable to demonstrate temporal patterns representing learning activities, and these patterns should be justifiable by theories. This finding aligns with previous studies that discussed the importance of theory in learning analytics (Gašević et al., 2017; Wise & Shaffer, 2015).

Chapter 5. The Framework

This chapter introduces a framework and addresses RQ3 of this thesis. Drawing upon the findings from the mapping study, I present a framework designed to guide temporal studies. The framework's structure and utility are detailed, explaining how it can be employed effectively, including a practical example for clarity. To demonstrate and evaluate its applicability in real-world scenarios, the subsequent two chapters will feature two follow-up case studies. These studies will serve both as illustrations of the framework in action and as evaluation of its effectiveness and practicality in research settings.

5.1. Proposing The Framework

RQ3: Based on the evidence uncovered, what framework can be developed to guide the selection of temporal techniques and learner data for deeper insights into the learning perspective?

To address this research question, we discuss the key components of temporal studies in the educational domain. Building upon this foundation, Owen and Baker (2020) proposed a framework comprising three main components: data collection, feature engineering, and data analysis. They suggested a framework for behavioral analysis in serious game design to systematically engineer features, producing meaningful results and providing insights into impactful interactions with the game, leading to users' learning. In other words, the framework enables researchers to systematically discover high-resolution features capturing fine-grained information about users' interactions and behaviors. It is important to note the necessity of aligning data collection and engineering with analytical techniques that enable high-resolution analysis. This process can be refined through critical and iterative interplay with research questions and available data (Owen & Baker, 2020).

We adapted and extended this framework to demonstrate how the findings of our mapping study can be utilized by researchers in the processes of data design, feature engineering, and analysis to implement a temporal model. In this framework (shown in Figure 5-1), we followed the general structure of Owen and Baker's original three components: Data Design and Collection, Feature Engineering, Analysis, and

Discussion. The elements within these components are shown as boxes. Additionally, we offer some recommendations for each component specific to temporal studies.

We modified the elements of Owen and Baker's model to adapt the framework for implementing a temporal model. Beginning with the Data Design and Collection component, the primary aim of this component is to ensure the collection and storage of all relevant data. Proper implementation of this component is crucial for the subsequent component, where the feature set will be derived from the collected data. However, it should be noted that in many studies, researchers relied on the underlying data collection framework of the system, such as an LMS, with all its benefits and limitations. Only when researchers had the ability to manipulate the data collection mechanism, as in the *Customized Log Files* category, could they fine-tune the raw data collection to be more directly usable by the selected analytical techniques to answer their research questions.

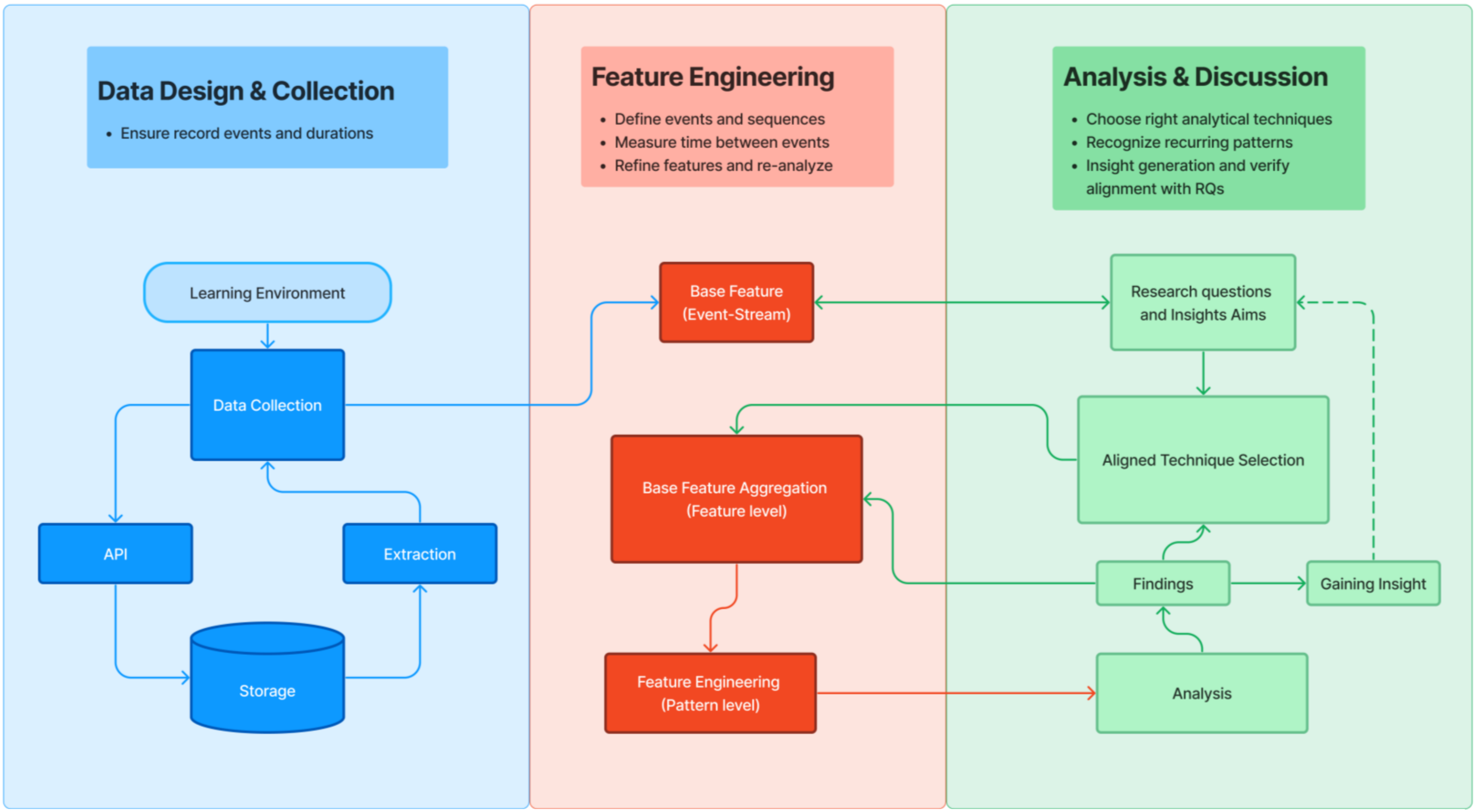

Figure 5-1. Our proposed framework to conduct temporal study.

In the middle component, Feature Engineering, the collected data are defined as an event stream that needs to be assessed through human judgment following the research questions. In other words, the defined events should carry sufficient information to allow for the analysis of the learning phenomenon. Assessing the event stream with a research question is, indeed, the first step of the analysis. Next, in the Analysis and Discussion component, the appropriate technique to address the research question is selected. After selecting the technique(s), appropriate features need to be engineered based on the technique requirements. This stage can be implemented through aggregating events in the event stream or by engineering more complex feature sets that represent patterns, such as creating a block of sequences of online activities (e.g., reading or watching a video). Returning to the Analysis and Discussion component, the engineered features are analyzed, and the findings are discussed. The results can be assessed, and, if necessary, iterative feature engineering can be performed to refine existing features or to design a new set of features to effectively facilitate the analysis stage. Finally, it is imperative that the findings lead to insights about learning, and researchers assess whether the insights constitute answers to the research questions.

We propose two starting points for this framework, depending on the purpose of its use. In the first scenario, which is applicable to most research analytics, the process can begin with a research question (RQ) and intended insight about learning. Researchers can rely on basic features and then follow the information flow of the framework. In this approach, one starts with an RQ, examines the base feature, and iteratively seeks to determine if the insights address the RQs. Later in this thesis, I will present two case studies that illustrate this scenario. I will discuss the extent of literature coverage for each case study and demonstrate how a rich, topic-specific literature base can provide a robust method for analysis. The alternative scenario if the goal is tool development, such as designing a new learning environment (akin to Owen and Baker's approach), the process might start from data design and collection. Researchers might conceptualize an analytical and monitoring system to evaluate if the tool is functioning as intended. Thus, the research process focuses on the extent to which the tool is monitored. In this scenario, the framework can guide the tool developers in how to develop the monitoring and analytical apparatus that can provide desirable insights into how the tool supports learning or examine where it does not meet the intended goals. In the second scenario, the research questions are formulated in response to learners' observed outcomes while using the tool or learning environment. Indeed, these are just two end-to-end example scenarios. Another possible scenario emerges where the framework serves as a resource for researchers interested in aspects of temporal analytics and their interconnections as found in the literature, (i.e. data features, techniques, insight, research questions).

5.2. Step-by-Step Example Within the Framework

To further clarify the framework's process, we provide an example of a research study that gained insight into learning through proper data design and collection, feature engineering, and analysis and discussion. Fan and colleagues aimed to explore online activities indicative of a self-regulated learning process (Fan et al., 2021a). Therefore, the targeted insight was identifying learning indicators. At the data design and collection stage, they used data from a flipped course encompassing a wide range of online preparatory activities, including reading an e-learning textbook, annotating notes, watching videos, participating in discussion forums, following up with quizzes, and other activities. They captured and cleaned all the activity log data (*raw data*) to create clickstream events (*feature-level data*). In fact, these events represent different activities (e.g., posting in discussion, watching videos). These events could potentially harbor SRL process indicators that address the study's research questions, implying whether analyzing the sequence of micro-level SRL processes can reveal distinct patterns in how often and in what order these processes are activated when students engage with different learning tactics. This could provide insights into how these tactics facilitate self-regulated learning, ultimately guiding more effective course design.

Next, based on their research questions, which required unveiling the temporality aspect of activities, they considered analytical techniques capable of revealing temporality. The researchers selected *frequent sequence analysis*, *clustering*, and *process mining* techniques for this purpose. Accordingly, based on their research questions and techniques, they engineered a set of features with temporality representation. Therefore, the study defined learning tactics and strategies – based on the SRL theory – representing students' approaches to learning the materials. Learning tactics are specific cognitive routines for tasks, while learning strategies are how a learner chooses and combines these tactics toward a learning goal. The tactics and strategies were defined by engineering the sequence of activities (e.g., watching a video, then working on an exercise, followed by taking a quiz, represented actions and sub-actions in SRL) in the given time period. These sequences of activities were, in fact, pattern-level features. Next, using the clustering technique, the study grouped these sequences and labeled them as different SRL tactics. Each tactic was discussed based on SRL sequence compositions. For example, Tactic 1 includes activities with a high frequency of reading materials along with taking quizzes.

Next, the authors explored transitions between the tactics to show students' engagement dynamics in terms of SRL processes. They interpreted changes in SRL processes as learning indicators. Finally, the authors elaborated on the findings and discussed how different groups of students approached learning materials differently, thereby deepening our understanding of the learning phenomenon.

5.3. Guiding Questions Through Using the Framework

Based on the finding from our mapping study, we propose a set of guiding questions for each stage of using the framework to conduct temporal study.

Data Collection

- What available raw data sources will you focus on: LMS log data, customized software log data, learning products, performance measures, etc.?
- Are specific features in the environment that could potentially impact learning indicators captured and included in the dataset?
- Are there aspects of the course design that impact on the learning indicators such as collaboration, feedback, and time on learning? If yes, can they be captured and included into the data?
- Are there learning products and performance measures being recorded and stored? Are there multimodal data sources like video records or eye-tracking?
- How frequently will data be collected? Is it in real-time or at set intervals?

Base Feature Extraction

- How will you define events based on the raw data?
- Will you be extracting time features, and if so, what specific time-related attributes will you focus on (e.g., lag time, time taken)?
- Are there any specific patterns in the data that you anticipate, such as event sequences or transitional patterns?
- Based on initial analysis, are there any other sequential patterns or transitional patterns that need to be considered?

Base Feature Aggregation (Feature level)

- How will you aggregate event data?
- Which specific traces (e.g., trace-exercise, trace-reading, trace-quiz) will be of primary interest?
- Are you interested in specific group of event patterns?
- Based on pattern recognition or unanticipated findings, do certain features need to be re-aggregated or refined?

Feature Engineering (Pattern level)

- How will you detect event sequences and transitional patterns?
- How is your selected technique limiting the choice of pattern features to be engineered?
- How will you differentiate between different types of sequential patterns, such as group event patterns versus other sequential patterns?
- How will summative features be calculated, such as averages or frequencies?
- After model testing, are there new patterns that have emerged that require additional feature engineering?

Research Questions/Aligned Technique Selection

- What is the primary focus of your research questions? (e.g., exploring socio-dynamics, method or algorithm development, at-risk student identification)
- Based on the research question focus, which analytical techniques will be most suitable? (e.g., process mining, cluster analysis, neural network)
- How will the chosen analytical technique align with the patterns and features identified in previous stages?
- After each iteration, have the results from the techniques used necessitated a revision or adjustment of the initial research questions?

Findings/Analysis

- What insights did you derive regarding the various learning indicators?
- How did the analytical techniques help in recognizing recurring patterns?
- How were visualization techniques employed to represent temporal patterns? Did they improve the comprehensibility and interpretability of the data?
- Were there any unexpected findings, and how do they align with the research questions?
- Did the features that you utilized yield interesting results and insight that you aimed for? If not, do you need to revisit earlier stages for further feature extraction, or engineering?

Gaining Insight

- How do your findings contribute to understanding the temporal aspects of learning?
- What recommendations or interventions can be derived from the insights?
- Are there any areas for improvement or further exploration based on the findings? Are there components in your framework that require revisiting?

5.4. Utilizing the Framework: A Guide for Researchers

Before delving into the utilization of the framework, I introduce the reference web tool[2]. This tool comprises a database of papers from the mapping study and is designed to have filters for the components of research categories discussed in the mapping study (Figure 5-2). Alongside the framework, this tool aids researchers in narrowing down and identifying similar studies.

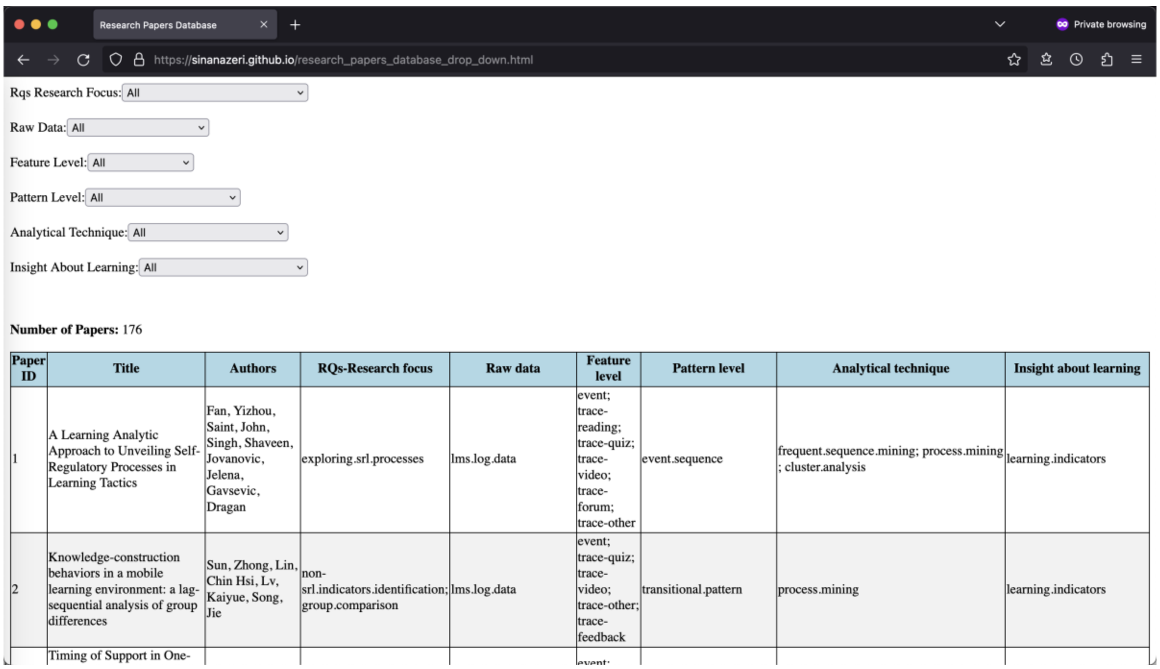

Figure 5-2. A snapshot of the reference web tool.

Before conducting research using our framework, the researcher can begin by reviewing the insights from existing temporal studies detailed in Table 4-11 and the reference webpage tool. This will aid in identifying the type of insights you aim to achieve.

Next, consider data collection and design. Assess the granularity and type of data that will support your research. Reflect on the questions we presented in the previous section regarding data collection and design. We have showcased common data types and their corresponding insights in Figures 15-17 for reference. The

---

[2] https://sinanazeri.github.io/research_reference_tool.html

references can be located by selecting the data criteria on our reference webpage. Choose relevant research question types and data that you have on the reference webpage.

After defining the direction of your study and its corresponding research questions, the next step is to choose the appropriate analytical techniques. We have highlighted various techniques that align with specific insights to streamline this process. Utilize our reference webpage to select the type of research question and insight to receive a list of relevant publications with similar research question aims and insights. From this list, you can identify the techniques utilized in the papers.

Transition to feature engineering. Create a novel set of features, particularly those that incorporate time features. These will be invaluable in revealing patterns in online behavior, especially if your research is based on LMS log data. Time features are also crucial in linking time dynamics to learning outcomes. Further feature engineering at the pattern level can also facilitate the analysis of temporal behavior. Our findings indicated that studies often proposed a method to model users' behavior by providing a temporal representation of the data that fits the technique. For instance, a main trend in revealing general characteristics in the sequence of actions is using process mining and finding associations between process mining results and assessment values (e.g., final grades), which can model students' online behavioral habits, i.e., sequences of actions, and interpreting these habits as learning indicators (e.g., Wang et al., 2019). Another trending research area in temporal studies is characterizing the block of online activities in a defined period. The most commonly used technique for this purpose is frequent sequence mining (FSM). In this case, it is also beneficial to cluster similar blocks of activities.

Lastly, it is important to note the incorporation of visualizations with other temporal techniques can provide extra explanatory power to the analysis. We found 62 studies (from our mapping studies) heavily relied on visualization to analyze their findings.

Chapter 6. Case study 1: Unveiling Temporality in Students’ SRL Processes

In this chapter, I demonstrate how utilizing the framework guided us in conducting a study to explore the aspect of time in students' Self-Regulated Learning (SRL) processes. Initially, I discuss the importance of the framework in the process of feature engineering and choosing the right analytical methods to gain a better understanding of learning dynamics.

We are building on previous research that identified different SRL phases in an online learning environment. This study aims to understand the dynamics of SRL phase transitions and to categorize different SRL processes based on these transitions. We are also interested in how these patterns correlate with students' academic performance. The insights from this study can contribute to implementing an automated feedback system.

Following that, I draw context from our previously published study by Nazeri, S., Hatala, M., & Salehian Kia, F. (2023). When to Intervene? Utilizing Two Facets of Temporality in Students’ SRL Processes in a Programming Course, published in the LAK23: 13th International Learning Analytics and Knowledge Conference (LAK2023). Association for Computing Machinery, New York, NY, USA, 293–302. The study can be accessed at https://doi.org/10.1145/3576050.3576095. Lastly, I discuss the findings and conclude this study.

6.1. The Use of the Framework to Conduct the SRL Study

Our framework contributed significantly to clarifying the steps necessary to conduct a temporal study. We adhered to the information process proposed in the framework to reveal the temporality in SRL behaviors in students. In this section, we demonstrate how the framework provided guidance for each step of conducting the study, as Figure 6-1 illustrates the different components of temporal studies and how to approach them. Since the data was already collected, we focused on the remaining two components: feature engineering, and analysis and discussion.

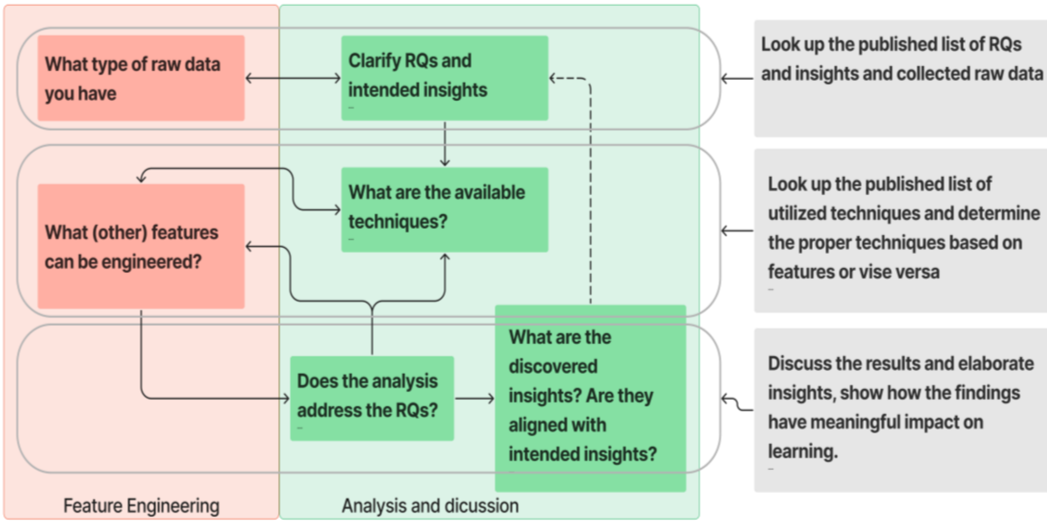

Figure 6-1. The framework process for conducting the temporal study.

6.1.1. Approaching RQs and Intended Insight Based on Available Data

The initial step involved clarifying the aims of the study with respect to the availability of data. Specifically, we needed to assess whether the data at hand were sufficient to address our study's aim of offering insights into learning and, if so, the nature of these insights. This study utilized data previously collected to understand the SRL phases during students' interactions with Learning Management Systems (LMS) (more details about the data in the method section). A preliminary review of Table 4-11 suggested that our objective could be to uncover learning indicators. From there, we considered whether our existing data sufficiently supported this goal and examined how our research focus might align with the categories outlined in Table 4-5. Therefore, based on the data we had, it became evident that our study's research questions (RQs) aligned with the group of studies in the category exploring SRL processes.

6.1.2. Exploring the Available Techniques and Feature Engineering

With a clear direction for our study, the framework aided us in identifying the right techniques and refining our feature set. By using our reference webpage tool, we were able to filter out relevant studies based on attributes that matched ours. We first shortlisted studies focusing on insights about learning indicators (n=77), then further narrowed it down to those exploring SRL processes (n=22). We then selected 15 papers that used LMS log data as their primary data source (the table below shows these result).

Table 6-1. List of relevant papers based on RQ focus, raw data, and aimed insight about learning.

ID	Title	Authors	Pattern level	Analytical technique
1	A Learning Analytic Approach to Unveiling Self-Regulatory Processes in Learning Tactics	Fan, Yizhou, Saint, John, Singh, Shaveen, Jovanovic, Jelena, Gavsevic, Dragan	event.sequence	frequent.sequence.mining; process.mining ; cluster.analysis
10	Theory-based learning analytics to explore student engagement patterns in a peer review activity	Er, Erkan, Villa-Torrano, Cristina, Dimitriadis, Yannis, Gasevic, Dragan, Bote-Lorenzo, Miguel L, Asensio-Perez, Juan I, Gomez-Sanchez, Eduardo, {Mart\'\inez Mones}, Alejandra	transitional.pattern	process.mining
20	Effects of a social regulation-based online learning framework on students' learning achievements and behaviors in mathematics	Hwang, Gwo-Jen, Wang, Sheng-Yuan, Lai, Chiu-Lin	transitional.pattern	process.mining; basic.statistical.analysis ; visualization.analysis
22	Process analysis of teachers' self-regulated learning patterns in technological pedagogical content knowledge development	Huang, Lingyun, Lajoie, Susanne P	event.sequence; transitional.pattern	cluster.analysis ; process.mining; visualization.analysis
28	Process mining for self-regulated learning assessment in e-learning	Cerezo, Rebeca, Bogarin, Alejandro, Esteban, Maria, Romero, Cristobal	transitional.pattern	process.mining; visualization.analysis
29	Trace-SRL: A Framework for Analysis of Microlevel Processes of Self-Regulated Learning From Trace Data	Saint, John, Whitelock-Wainwright, Alexander, Gasevic, Dragan, Pardo, Abelardo	event.sequence; group.event.pattern ; transitional.pattern	process.mining; cluster.analysis
43	How Patterns of Students Dashboard Use Are Related to Their Achievement and Self-Regulatory Engagement	Kia, Fatemeh Salehian, Teasley, Stephanie D, Hatala, Marek, Karabenick, Stuart A, Kay, Matthew	event.sequence; group.event.pattern	frequent.sequence.mining; cluster.analysis ; visualization.analysis

55	The importance and meaning of session behaviour in a MOOC	de Barba, Paula G, Malekian, Donia, Oliveira, Eduardo A, Bailey, James, Ryan, Tracii, Kennedy, Gregor	summative	cluster.analysis ; visualization.analysis
65	Exploring sequences of learner activities in relation to self-regulated learning in a massive open online course	Wong, Jacqueline, Khalil, Mohammad, Baars, Martine, de Koning, Bjorn B, Paas, Fred	event.sequence; transitional.pattern	frequent.sequence.minin g;process.mining; visualization.analysis
66	Analytics of Learning Strategies: Associations with Academic Performance and Feedback	Matcha, Wannisa, Gavsevic, Dragan, Uzir, Nora'Ayu Ahmad, Jovanovic, Jelena, Pardo, Abelardo	event.sequence; group.event.pattern ; transitional.pattern	frequent.sequence.minin g; process.mining ; cluster.analysis ; visualization.analysis
67	Analytics of Learning Strategies: The Association with the Personality Traits	Matcha, Wannisa, Gavsevic, Dragan, Jovanovic, Jelena, Uzir, Nora'ayu Ahmad, Oliver, Chris W, Murray, Andrew, Gasevic, Danijela	event.sequence; group.event.pattern	frequent.sequence.minin g; cluster.analysis
82	How social challenges affect children's regulation and assignment quality in hypermedia: a process mining study	Paans, Cindy, Onan, Erdem, Molenaar, Inge, Verhoeven, Ludo, Segers, Eliane	transitional.pattern	process.mining; visualization.analysis
138	Learning analytics to unveil learning strategies in a flipped classroom	Jovanovic, Jelena, Gavsevic, Dragan, Dawson, Shane, Pardo, Abelardo, Mirriahi, Negin	event.sequence; group.event.pattern	frequent.sequence.minin g; cluster.analysis ; visualization.analysis
165	Understanding student learning pathways in traditional online history courses: utilizing process mining analysis on clickstream data	Crosslin, Matt, Breuer, Kimberly, Milikic, Nikola, Dellinger, Justin T.	event.sequence; transitional.pattern	process.mining; frequent.sequence.minin g; cluster.analysis
173	Diagnosing virtual patients in a technology-rich learning environment: a sequential Mining of Students' efficiency and behavioral patterns	Zheng, Juan, Li, Shan, Lajoie, Susanne P.	event.sequence; group.event.pattern	frequent.sequence.minin g; cluster.analysis

This phase helped us to identify potential candidates for pattern-level feature engineering and selecting the right techniques for our study. To further apply or derive inspiration from the information in the selected papers, we utilized guiding questions from the previous chapter, which helped refine our feature engineering and technique choices. For instance, from Table 6-1, we selected four papers closely aligning with our study (Fan et al., 2021a; Jovanović et al., 2017a; Kia et al., 2020; Matcha, Gašević, Uzir, et al., 2019a). We then explored the pattern-level features they employed, as detailed in Table 6-2. In considering features, we contemplated questions such as:

- *How will you define events based on the raw data?*
  Our primary features were derived from the LMS interaction clickstream. Preparing this data required data cleaning and validation prior to conducting analysis (details are provided in the method section of this chapter).

- *Which specific traces (e.g., trace-exercise, trace-reading, trace-quiz) will be of primary interest?*
  We were particularly interested in specific types of student interactions, such as those with guiding pages about course modules and exercises. By analyzing the sequences in which students visited different pages, we could infer aspects of their SRL behavior. In a previous study, we traced various activities to develop indicators for different SRL phases. For instance, the pattern of a student viewing the Organizer page, then the Overview page, followed by the Organizer page again, and finally the Specification page, can suggest the phase of task definition in SRL. Figure 6-2 in this chapter illustrates how we detected these phases through the sequences of interactions with content pages in the LMS (more detail in the background section of this chapter).

- *Will you be extracting time features, and if so, what specific time-related attributes will you focus on (e.g., lag time, time taken)?*
  The study by Fan et al., (2021) highlighted the importance of recording the gap

time between activities, noting that a lack of data logging for 45 minutes indicates the learner has abandoned the learning task, thus defining the end of a learning session. Similarly, another paper (Jovanović et al., 2017a) identified a 30-minute threshold of continuous learning activities as preparation for face-to-face classes. We also considered time-related attributes, such as the duration and start time of each page visit. This analysis allowed us to investigate the timing and duration of sessions of different learning activities as well as SRL phases.

After studying works identified by our proposed framework (Table 6-2), techniques including frequent sequence mining, clustering, and visualization were used most often among the candidate studies.

Table 6-2. Candidates that are highly aligned with our study.

ID	Pattern-level Feature Engineering	Applicability to Our Study
1	Event sequence: Learning session defined as a 45-minute cut-off for time window of learning activities consisting of coding exercises, watching videos, readings, assignments, annotation, forum discussion, and visiting information webpage. The sequence of these activities was represented in a session. Group event sequence to identify SRL sequential behavior. Transitional pattern analysis of students moving between different sequential behaviors.	Session of activities within an appropriately fixed time window or cut-off time value. Clustering the sessions to identify distinct patterns.

43	Event sequence: 10-minute cut-off time window of log data dashboard view, determined based on a typical 5-minute user session. Group event sequence: similar pattern of close consecutive events within window represented as (e1,e2), and two subsequent sessions with elapsed time greater than 10 minutes are represented as (e1, e2) − (e3).	Windows of appropriate size with a cut-off value. We used a 30-minute cut-off determined based on statistical analysis of gaps between SRL phases.
66	Event sequence: Sequence of activities within a 21-minute interval, based on quartile analysis. Group event sequence: group similar sequences to identify learning tactics. Transitional pattern: analysis of students moving between different tactics (group of sequences)	One way to identify the cut-off value for creating a session is using median and researcher intuition.
138	Event sequence: Learning session defined as a 30-minute time window for consecutive learning activities, represented as a sequence of activities within the session. Group event sequence: Group event sequence to identify similar patterns in sessions.	Session of activities. We used sessions of SRL phases. Categorizing similar sessions.

6.1.3. Analyzing Results and Refining Further RQs and Feature Engineering

Based on our chosen techniques and data, our findings for the SRL study highlighted various SRL behaviors among students. Our study has similarities to studies listed in Table 6-1. Addressing the guiding questions can help us to assess and refine the findings:

- *How do your findings contribute to understanding the temporal aspects of learning?*
  Our results illuminate the temporal dynamics of learning. We identified and categorized SRL processes and further analyzed how these processes correlated with assignment grades.

A notable aspect of our framework is its iterative nature, which encourages continuous evaluation and refinement of our results and insights. This iterative approach fosters innovation and a deeper understanding. Addressing the following guiding questions (taken from the previous chapter) would help us in this respect.

- *What recommendations or interventions can be derived from the insights?*
  Our identification of diverse SRL processes can guide interventions to improve students' SRL strategies. The timing of such interventions is vital; hence, discerning the right moment to step in is important. **We realized that less attention has been given to this aspect**. Table 6-3 provide a summary of how the study can be improved by considering this aspect.

- *Did the features that you utilized yield interesting results and insight that you aimed for? If not, do you need to revisit earlier stages for further feature extraction, or engineering?*
  In the first iteration, we uncovered patterns in SRL phase transitions. However, we were also interested in the timing aspects of those transitions. Therefore,

more feature engineering is required to capture patterns in time intervals between SRL phase transitions.

- *Are there any areas for improvement or further exploration based on the findings?*
  Are there components in our framework that require revisiting? To delve deeper into the time intervals of students' SRL activities, we should consider developing a new set of features representing the timing of SRL phases.

Table 6-3. Proposing improvement: Identifying Optimal Intervention Timings

What is Missing	**New Features to Consider**	**Appropriate Technique**
Understanding optimal timing for SRL interventions	Calculating timing and transitions between SRL phases and SRL processes.	Descriptive statistical analysis and visualization of key transition phase timings.

In summary, our framework guided us in conducting temporal research. It not only directed us in selecting appropriate analytical techniques and feature engineering methods but also prompted us to approach the study with an additional aspect of temporality.

6.2. Introduction and Research Questions

Learning programming has become a vital skill in many different professions, even beyond the computer science field. Yet, acquiring programming skills poses many challenges to students who are required to learn a wide range of skills (Hertz, 2010). To succeed in programming, students need to master problem-solving skills, which demand understanding the tasks, dividing them into subtasks, solving the subtasks, gaining reflection from the outcomes, and iterating the process (Fitz Gibbon et al., 2010; Kuo & Hsu, 2020). Mastering these skills and becoming successful in the programming course overlap with higher-level general skills of self-regulated learning (SRL). SRL is a conceptual framework that encompasses cognitive, meta-cognitive, motivational, and emotional aspects of learning (Zimmerman, 1990). It is an iterative process that includes a set of actions for familiarizing with the task, planning to solve them, engaging with the task, and self-evaluation. There is a strong body of evidence that learners who possess SRL skills achieve a higher success rate in their learning and showed higher effective learning outcomes. For example, the studies showed that the SRL process contributes to self-awareness and motivational aspects of an individual to trigger meta-cognitive activities to acquire information (P. Winne & Hadwin, 1998); a general study by Zimmerman showed that SRL is critically related to academic performance because it is defined as a key component of learning (Zimmerman, 1990). In recent years, the research on SRL has expanded to studying how it affects learning in many disciplines, including in the programming context. For instance, Dominguez and colleagues developed a learning tool to facilitate students' SRL activities to learn a programming language (Dominguez et al., 2021); another study found that limited SRL skills are associated with more programming errors (Loksa & Ko, 2016); similarly, a study investigated how successful students develop SRL skills in a programming course (Falkner et al., 2014). These studies went beyond survey data, a common tool in SRL research, and utilized clickstream log data (Bernacki, 2018).

Identifying SRL-associated patterns of actions from the clickstream log has become a dominant approach to understanding SRL processes in the learning environments in use today. The primary goal of this study is to provide a better understanding of SRL processes used in programming problem-solving tasks and identify when the support for improving the SRL ability of the students can be deployed. Ultimately, our goal is to develop an automated detection of students' SRL processes so they can inform either teacher or an automated agent about the timing of suitable interventions. To identify patterns in SRL-associated actions, first, the actions need to be theoretically grounded. Although many studies utilized clickstream data and theorized certain combinations of actions in LMS are associated with SRL processes, further study is needed to understand to what extent students think about their actions as being part of their intentional learning regulation and, in turn, represent steps in their SRL process. In this study, we utilized the data from the previous study (Salehian Kia et al., 2021), where actions representing SRL phases were grounded in the SRL theory proposed by Winne and Hadwin (Winne & Hadwin, 1998). In the study (Salehian Kia et al., 2021), we utilized a cross-validated approach using log data and a self-reported survey approach to identify and verify the identification of SRL phases as students worked on programming assignments. In the follow-up study, we showed how SRL processes, the sequences of SRL phases, categorized into four SRL process types using a theoretical-pragmatic approach and coded by human experts, interact with the level of domain knowledge and explain a significant amount of variance in students' assignment grades (Hatala et al., 2023a). In the present study, we extend our approach to computationally identify the patterns that are associated with SRL processes. One of the contributions of this study is to examine to what extent automated detection of patterns of SRL processes is accurate when compared to the expert coders. In other words, this study aims to propose how feasible it is to develop an automated detection of the SRL processes that can lead to implementing a feedback/intervention system.

Secondly, this study also investigates the temporal nature of the SRL processes. The temporal aspect of SRL can be conceptualized as having two facets (Knight, Friend Wise, et al., 2017): temporal as a series of actions in a particular order, and temporal as measuring the time instance when a particular action happens. The first facet has been extensively focused on in SRL studies in recent years (e.g. (Cheng et al., 2017; Fan et al., 2021b; Saint et al., 2021; Siadaty et al., 2016)). However, less attention was given to the second facet. This study's contribution is in this area by investigating timing patterns within the timespan of SRL processes and how consistently, from the timing perspective, students behave with respect to their SRL phase transitions. This can lead to implementing an automated system to monitor students' behavior and intervene to improve SRL processes in a timely manner.

To fulfill the aforementioned goals, we proposed the following research questions:

RQ1. What are the students' SRL behavioral patterns during the two weeks of work on the assignments?

RQ2. How do the computationally detected SRL process patterns agree with human expert identification?

RQ3. What are the associations between the identified patterns and academic performance?

RQ4. Can we identify the temporal dynamics in the SRL phases transitions across students?

6.3. Background

Self-regulated learning

SRL is often a cyclical process through different phases (Panadero, 2017). Although named differently, Zimmerman's (2002) and Winne and Hadwin's (1998) models appear to agree on the three essential phases of preparation, execution, and reflection. In the preparation phase, students gain an understanding of the task, set goals, and plan how to approach the task. They work towards the goal in the execution phase by engaging with the task. In the reflection phase, they evaluate and possibly adapt their approach. While SRL processes are cyclical in nature, learners can jump between SRL actions, including the iterative execution of phases (Molenaar & Järvelä, 2014). The models theorize how students transition from phase to phase when they are accomplished, self-regulated learners. However, students vary in their level of SRL skills; these skills are either self-learned or obtained through SRL-specific training (Theobald, 2021). With the recent broad availability of trace data from learning environments, a significant body of research aims to detect SRL actions, phases, and processes (Bernacki, 2018), to support learners in improving SRL skills through SRL interventions, such as via prompts embedded in the learning environments (Ifenthaler, 2012; Moos, 2017). Although these approaches are successful in the SRL detection phase (e.g., (Saint, Gašević, et al., 2020; Salehian Kia et al., 2021)), many open challenges have to be solved before the technology can be widely deployed (Azevedo et al., 2017; Moos, 2017). Understanding the temporal aspect of SRL, an essential aspect for timing the intervention, is particularly challenging due to task dependency and other contextual factors (Bernacki, 2018; Molenaar & Järvelä, 2014). This study contributes to understanding the temporality of students' unfolding SRL process in the 2-week-long programming problem-solving assignments.

6.3.1. Dynamic and Temporal Aspects of SRL and phases

The cyclical nature of the SRL unfolds over time as a process that is influenced by the learning task context (Roll & Winne, 2015). Molenaar and Jarvela (2014) identified two distinct strands of research on the temporality of SRL: as a relative arrangement among multiple events and as a continuous flow of events. The relative arrangement of events has been the main focus of many recent studies and falls into the category also identified by (Knight, Friend Wise, et al., 2017) as a series of actions. This body of research focused on identifying frequent *sub-sequences of activities* in the learning environment theorized to represent SRL events; the sub-sequences were typically labeled as strategies. In the next step, students with a similar composition of sequences, often determined as counts of strategy use, are clustered, and differences between clusters are examined. An example of such research is examining learning strategies by clustering sequences of activities in the flipped classroom setting (Jovanović et al., 2017b). The study found five different adopted strategies, and the authors discussed how the students tend to change the strategy to adopt a more effective one. In our previous work (Salehian Kia et al., 2021), we have defined indicators of SRL phases, such as task definition, planning, enactment, and adapting (see below for details), by a close reading of the student log files using a technique called text replay tagging (Baker et al., 2006). These indicators were specific to the exact instructional design template designed to facilitate SRL detection. The indicators have proven to have relatively high accuracy in detecting SRL phases (kappa = 0.68-0.72), even when deployed in the context of a different course (Salehian Kia, 2021).

The second group of studies examining temporality as a relative arrangement of events focuses on the SRL process and its structural properties. The prior research showed that successful students' processes are close to the theoretical SRL models (Bannert et al., 2014; Kokoç et al., 2021). One typical approach is exemplified by the study (Saint, Whitelock-Wainwright, et al., 2020a), where processes are derived using the process mining technique. In the next step, students' processes are clustered using a similarity metric, such as the similarity distance between measures in hierarchical clustering, and the clusters of students are compared. An opposite approach is to define the structural properties of the SRL process from the theoretical perspective, such as the presence and the level of iterations over SRL phases, to represent the level of maturity of students' SRL skills, and then use these characteristics to categorize students' processes. We have used such an approach in our previous study (Hatala et al., 2023a), where we detected SRL phases using indicators from Salehian Kia et al., (2021). We have found high stability in students' SRL processes across the five consecutive 2-week assignments. Additionally, we found that the SRL process level had the highest impact on students' assignment grades for those with the lowest programming proficiency.

The type of SRL research that falls into the second strand of research on temporality identified by Molenaar and Jarvela (2014) looks at individual temporal characteristics, such as positioning, duration, and rate. Knight et al. (2017) are more specific in delimiting this direction of inquiry on temporality, focusing on the time instance and timespan of the events. The research in this area is highly contextually dependent on the task and students' environment at different levels, as defined by (Ben-Eliyahu & Bernacki, 2015). Although generic guidelines for incorporating prompts exist, such as those outlined in Moos (2017), we have not found research that would analyze the specific timing of SRL events to support the timing of the interventions as students learn in LMS. The research investigating the timing of SRL events and time for intervention focused more on Advanced Learning Technologies, such as Intelligent Tutoring Systems, which have much tighter control over the unfolding learning process, and some are built explicitly on the SRL theoretical framework (Azevedo et al., 2017). Reviewing this body of work is not relevant to this paper. We focus on studying temporal aspects of the processes detected from clickstream data originating from LMS, an open learning environment where students engage in any manner that suits their learning.

6.3.2. SRL processes in programming problem solving

The importance of SRL has been highlighted in the context of programming problem-solving (Dominguez et al., 2021; Hatala et al., 2023a; Loksa & Ko, 2016) where the learner utilized its cognitive and metacognitive skills to attain the goals (Azevedo et al., 2010; Panadero, 2017; P. H. Winne, 2010). In a mixed-method study, Falkner et al. showed that students used a wide range of SRL activities relevant to the development process, problem decomposition, time management, and assessment difficulties to approach structured programming exercises (Falkner et al., 2014). Then, these activities were identified through qualitative coding analysis and mapped into SRL phases provided by the Zimmerman model (Zimmerman, 1990). The results showed that the majority of activities were associated with the goal-setting and planning phase (29%). A similar follow-up study showed the importance of discipline-specific design to improve SRL skills which led to success in programming skills development (Falkner et al., 2015). They observed how the deployed strategies and tactics of first-year students had different rates of final-year students. Another study explored the effect of explicit guidance and asserted that educators should consider the SRL skills of the students especially paying attention to those with low SRL skills (Loksa et al., 2016). The authors examined 31 undergrad students during a 10-week programming course. They coded both programming behaviors and SRL behaviors, and they found that the most common behavior was *process monitoring,* and the least common behavior was *adapting a solution*. Overall, the aforementioned studies highlighted the importance of the SRL process in developing programming skills, and identifying SRL processes is the key to providing constructive feedback/intervention to improve students' programming behavior in their subsequent assignments.

6.3.3. Detection of SRL-associated activities

In line with our previous study (Salehian Kia et al., 2021), to identify SRL-associated activities, we utilized the well-established SRL theory from the Winne and Hadwin model (Winne & Hadwin, 1998). This model fits properly in the material-rich learning systems, and it emphasized study as a goal-oriented activity through four phases, including task definition, planning, enactment, and adapting (see more details below). In our previous study, we developed a set of indicators of SRL phases. Figure 6-2 provides the scheme for phase detection through sequences of interactions with information elements, i.e., content pages, in the LMS. By developing a specific instructional scaffold for the assignments, we have imbued the pages within the scaffold with specific SRL-related semantics. For example, the sequence of viewing the page Organizer (our code of the top assignment scaffolded page), viewing the Overview page, viewing again Organizer, and viewing Spec (our code for the assignment detail specification page) is an indicator of the SRL phase of the task definition (see Figure 6-2). We verified the indicators by comparing the detected SRL phases from the LMS log files with self-reported student-identified phases collected just in time as they worked on the assignments. The resulting accuracy of indicators to detect SRL phases measured by weighted kappa was between 0.68 and 0.74. More details can be found in (Salehian Kia, 2021; Salehian Kia et al., 2021). As we use the same apparatus to detect phases in this study, we provide a pertinent summary of phases and indicators below:

Task definition (referred to as D further): the first phase implies that the student gains an understanding of the task and its aspects. In our scaffolded environment, several pages are designed to inform students about the overall task and available resources, help to estimate time for developing solutions, and understand the assessment of their work.

Planning (P) and goal setting: the second phase of the SRL study process usually involves setting or reassessing the goals and standards for the student's learning. In this phase, students set, or compare with previous, concepts of the task with personal standards to build motivational orientation. They plan how to approach a learning task, possibly decompose problems into sub-task, gain an understanding of the task components, compare subtasks, seek information, set time, and develop tests. The indicators of the planning phase (P) identify when the student access pages concerning task operations and objectives.

Enactment (E): the learner employs strategies and tactics which were planned in the earlier stage. While two previous phases were carried out as students interacted predominantly with the information contained in the LMS, the students' work in the enactment phase (and the adapting phase) switches between the external programming environment (not tracked) and the LMS to work with the details of the task specification. The indicators of this phase are students interacting with the specific limited set of learning materials.

Adapting (A): learners in this phase evaluate their progress on subtasks against the plan and requirements. Similar to the enactment phases, this involves alternating between the external programming environment and accessing information in LMS that facilitates detecting and correcting misalignments between goals and current work products with respect to set standards.

This study also builds on part of our second study (Hatala et al., 2023a), in which we have defined SRL process types from the theoretical-pragmatic perspective and manually coded the sequences of SRL phases detected using the indicators described above. In this study, we compare to what extent automatically detected SRL process clusters share their process structural characteristics with the theoretically defined types and what level of agreement exists between the two classifications. Here we present shortened definitions of SRL processes from (Hatala et al., 2022):

Type T0. SRL processes of this type are incomplete with respect to phases E and A. Although these processes tend to be short, such as D-P (i.e., phase P follows phase D) a longer process in which the same phase code is repeated in the subsequent sessions, but does not demonstrate the student's engagement of various metacognitive phases, is also considered incomplete, e.g., D-D-E-E-E-E. Iterations are not present in this process.

Type T1. A straight process through SRL phases without iteration(s).

Type T2. SRL process with iterations in the execution phase, i.e., iterations between phases E and A. The main characteristic of this process is that it demonstrates students' ability to monitor their progress in the execution phase and re-engage with the task. The iteration over alternating A and E phases can be repeated several times.

Type T3. SRL process with the 'complete-cycle' iterations, i.e., includes iteration through the planning phase. Such a process demonstrates the student's ability to partition the task into sub-tasks of smaller complexity, addressed iteratively. The iteration through the E and A phases at the sub-task level can also be present.

6.4. Method

6.4.1. Participants and their SRL processes

The data utilized in this study include 54 students, 18 to 24 years old, in five 2-week assignments in a university programming course. The programming tasks were presented through the scaffolded design, which allowed us to trace students and interpret interactions with the LMS from the perspective of SRL (Salehian Kia et al., 2021). We set a 20-min timeframe as an activity session to partition the log data, which resulted in 21,295 sessions. After eliminating those sessions that did not contain SRL phase indicators (Figure 6-2), 4,307 sessions remained. Next, we identified the SRL-associated phase for each session and generated the sequence of phases for each student in each assignment.

A sequence of student's sessions in an assignment, each session classified as one of four SRL phases, represents an SRL process. For example, the sequence of sessions D (task definition) → P (planning) → P → E (enactment) → E → E → A (adapting) → E → E → A is an SRL process consisting of 10 SRL phases. We generated 260 (54*5) sequences of SRL phases, each sequence representing one SRL process. This study aimed to investigate patterns within these processes. Next, we truncated similar consecutive phases into a single phase because this study is interested in the conceptual transitions of what students are doing from the SRL perspective. In the above example, the process showed several enactment phases (e.g., E → E → E); conceptually, this means that the student was working on the programming task in three sessions, enacting their planned strategies and tactics. It is the act of switching to a different SRL phase that conveys a student is self-regulating their learning. Therefore, in this study, we consider E → E → E → A to have a conceptually similar meaning to E → A.

6.4.2. Data analysis

This study is interested in how students approach their learning while working on the assignments from the timeline perspective; detecting patterns in SRL processes is key to revealing their approach. We utilized the clustering technique to distinguish different categories of the processes (addressing RQ1). The technique has been widely used in the body of literature (Jovanović et al., 2017b; Saint, Gašević, et al., 2020;

Siadaty et al., 2016). To address RQ2, we utilized the data from our previous study, where we qualitatively distinguished different types of SRL processes (Hatala et al., 2023a). We assess the agreement between the output from the clustering technique and the qualitatively classified SRL processes by examining the confusion table and computing Cohen's kappa. This stage shows the reliability of computationally detected patterns to implement an automated feedback system. Next, to address RQ3, we utilized the ANOVA test to determine the association of the SRL process types, as determined by clustering, with academic performance.

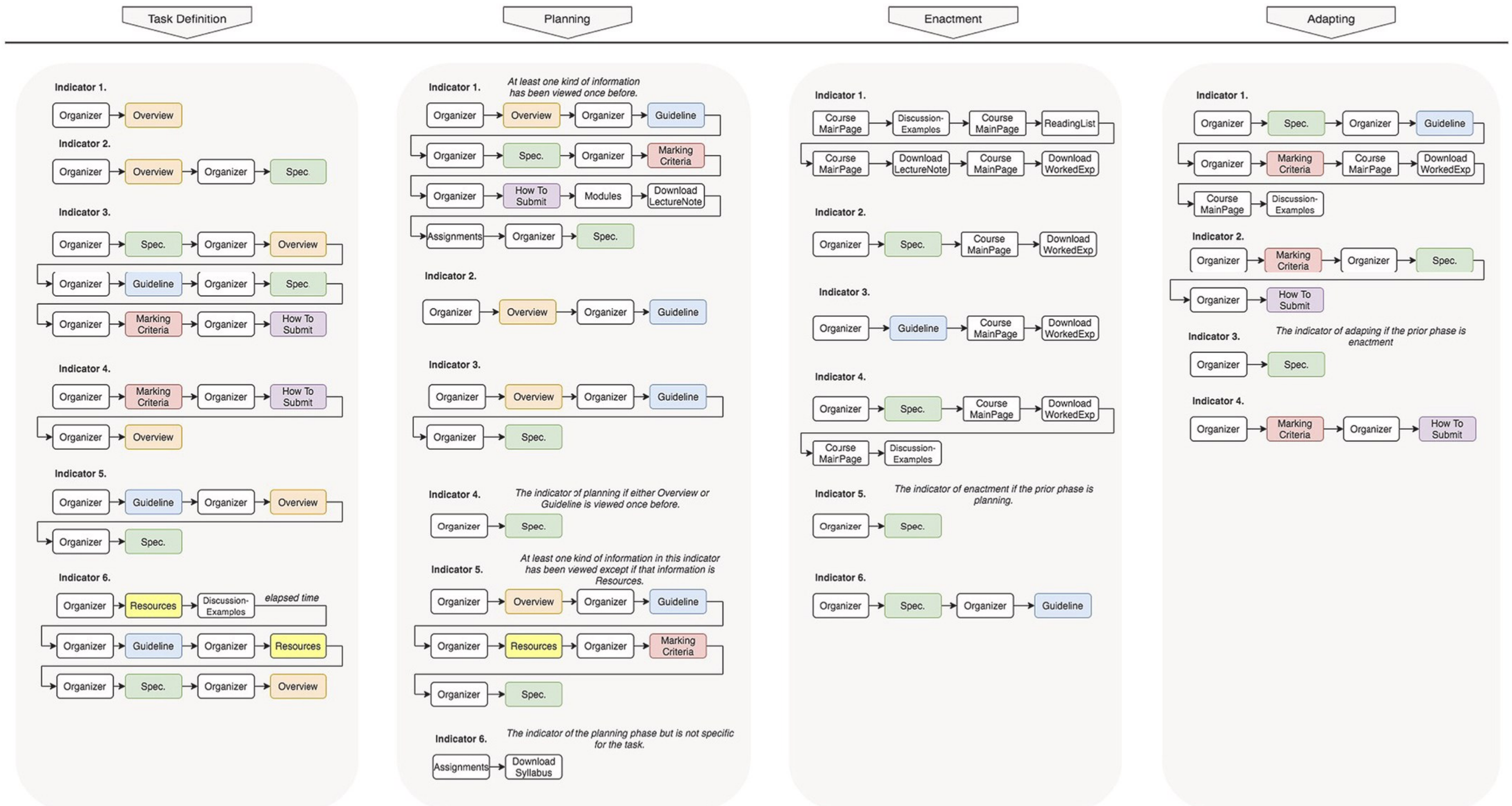

Figure 6-2. The indicators of SRL phases driven from students' log data.

6.4.3. SRL phase transitions temporal analysis

We are interested to understand if there is sufficient similarity in the time profile of the SRL phase transitions, at the SRL process type level, to determine an appropriate time for an instructor or a system to intervene. For this purpose, we analyze the timing of the SRL phases, both in their duration and, more specifically, time gap profiles between sessions where one SRL phase transitions to another. This provides insights into how students behave during the two-week assignment; are their SRL temporal behaviors similar? Are there any patterns in terms of how long certain phases are (e.g., the transition from E → A)? To address RQ4, we calculated the gap time between phases of SRL processes for every student in each assignment. This calculation was conducted before truncating the phases – removing similar consecutive phases (see section 3.1). Therefore, we create a gap profile for each individual during each assignment (e.g., D → 3 min → D → 55 min → P → 4 min → P → 110 min → E). Then, we investigate the distribution of these gaps for each phase pair of phase transition (e.g., D→P, E→A).

6.5. Results

6.5.1. RQ1: What are the students' SRL behavioral patterns during the two weeks of work on the assignments?

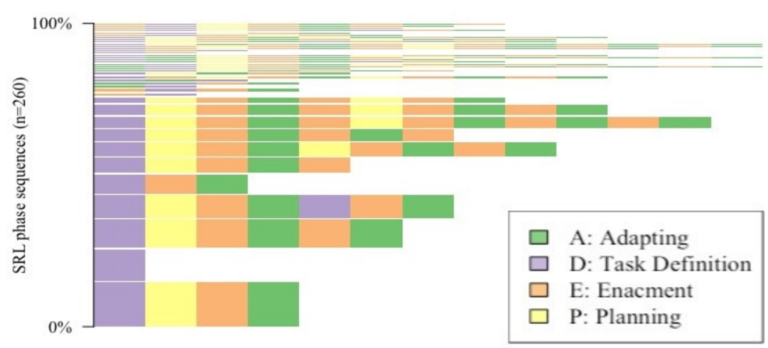

Figure 6-3. All SRL process.

Figure 6-3 shows the total generated SRL processes that contain the sequence of phases (n=260). As can be seen, the SRL sequences (i.e., SRL processes) vary. We utilized the agglomerative hierarchical clustering technique to differentiate the patterns among the sequences. To determine the optimum number of clusters, we utilized Silhouette and the gap statistic techniques that resulted in suggesting four clusters; pictured in Figure 6-4. The clusters indicated structurally different SRL processes adopted by students during five 2-week assignments. The clusters that contribute to our first RQ can be theoretically characterized as follows:

Incomplete SRL process cluster (CL 1, n=39). This cluster has the smallest number of SRL processes, and it represents incomplete SRL processes. This cluster is dominated by only task definition phases and includes a few other incomplete SRL phases. The absence of planning is apparent in this cluster.

SRL processes without SRL iteration (CL 2, n=75). The second most populated cluster comprises mainly complete SRL processes without iterations. However, processes missing one of the four phases comprise 45% (n=34) of this cluster. These incomplete processes mainly lack the planning phase (n=21) or task definition phase (n=7).

SRL processes with the iteration of enactment and adapting phases (CL 3, n=97). This cluster holds the greatest number of SRL processes (37.4%). The processes often start with task definition (D) and planning (P). In some sequences, one of D or P might be missing due to the potential conflation in the identification of these two phases within a single session. The main characteristic of this cluster is the presence of two enactment phases in each sequence, the second usually following the adapting phase. This can represent the demonstration of the student's self-monitoring activity. Therefore, the iteration of E→A (enactment followed by adapting) can be a desirable transition that programmers should master and demonstrate as they learn to program.

SRL processes with the iterations of the planning phase (CL 4, n=48). This cluster contains the lengthiest sequences, and it is among the least populated cluster (18.5%) after CL 1 (15.1%). The main characteristic of this cluster is a complete 'ideal' SRL cycle followed by the planning phase and continues with the iteration of enactment and adapting. Sequences in this cluster represent students' ability to plan their learning in sizable chunks, and adjust or expand their plans to keep the learning manageable. This is also ideal from the programming perspective as the strategy of dividing the main task into subtasks is one of the key programming strategies to manage complexity.

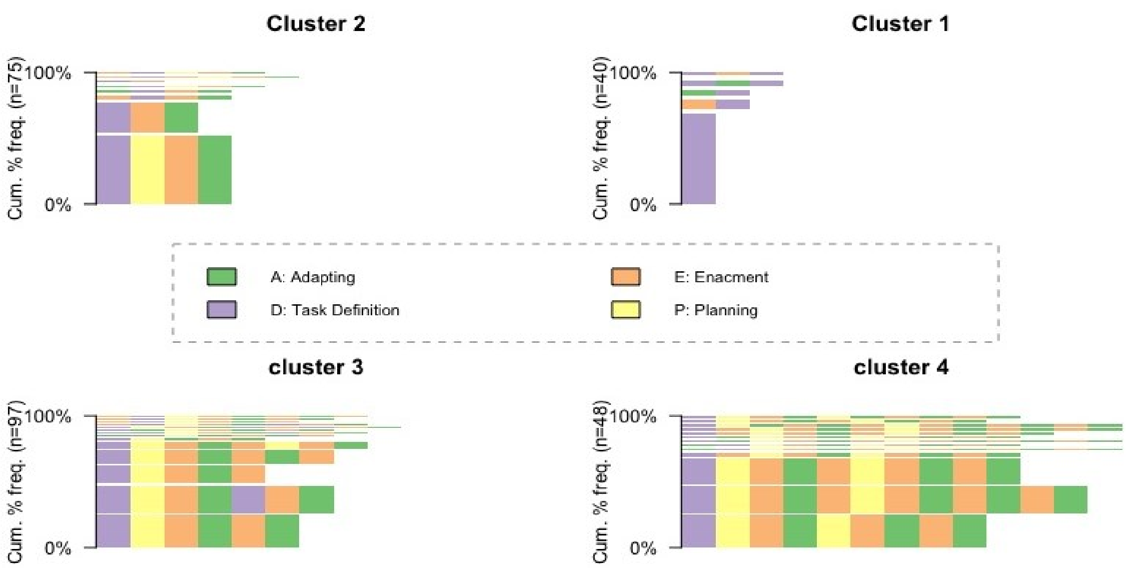

Figure 6-4. SRL process with respect to clusters.

6.5.2. RQ2: How do the computationally detected patterns agree with human expert identification?

In our previous study, we defined four distinct SRL process types from the theoretically-pragmatic perspective (see Section 2.4) and showed how effective they are when students use them (Hatala et al., 2023a). The sequences of SRL phases, the same as those used in this study, were then coded by human coders who achieved a high level of agreement.

In this study, we examined how computationally detected clusters are similar to human judgment. First, in both methods, four groups of SRL processes were identified. Second, the characteristics of the clusters were similar in both models, with an agreement of 85% and weighted Cohen's kappa= 0.84. Table 6-4 is the confusion table that shows the identified four types of SRL process coded by the experts (T0, T1, T2, and T3) against four clusters from the hierarchical clustering method (CL1, CL2, CL3, and CL4). The main disagreement happened in CL 3, where the algorithm clustered 33 processes with task definition or planning phase iterations to cluster 3. The main reason is that the clustering technique is more sensitive to the length of sequences rather than their meaning. Overall, our findings suggest that the clustering technique showed an acceptable performance to characterize the sequences.

Table 6-4. the agreement on the SRL process between the human coders (SRL Type(T)) and computer identification (SRL cluster (CL)).

Types vs Clusters	T0	T1	T2	T3
Cl1	39	0	0	0
Cl2	1	73	1	1
Cl3	0	1	63	33
Cl4	0	0	8	40

6.5.3. RQ3: What are the associations between the identified SRL process patterns and academic performance?

Using the ANOVA test, we found significant associations between SRL process clusters and the assignment grades $F(3,255)=10.47$, $p<0.001$. Further, Tukey HSD test for pairwise comparison of the clusters with respect to the assignments' grades showed significant differences in grade means for the cluster pairs CL1-CL3, CL1-CL4, and CL2-CL4 (Figure 6-5). In other words, the grades in assignments where students followed the SRL process classified to cluster 4 (CL4) were significantly higher than the first two

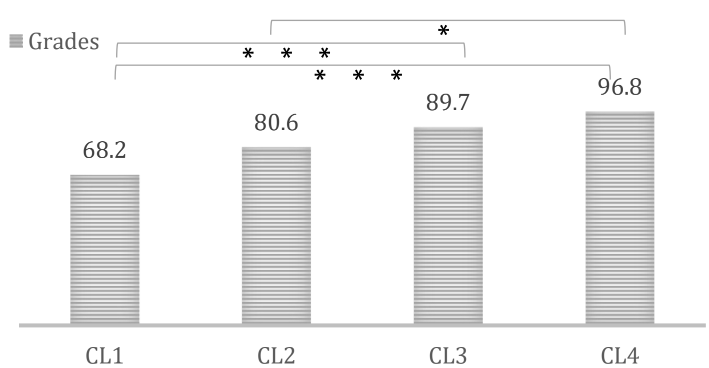

Figure 6-5. The mean of grades for each cluster. The asterisks (*) illustrates the significance (p<0.05) of difference, and three asterisks represents p<0.001.

clusters, which had SRL processes missing on some cyclical aspects of the SRL. The highest mean differences were between CL1-CL4 (28.6%) followed by CL1-CL3 (21.5%). The lowest mean differences are between CL4-CL2 (16.2%). Overall, our findings suggest that students in the two clusters with complete cycles and further iterations in phases seem to achieve higher grades. However, further studies are needed to clarify the particular qualitative details in SRL phase iterations that contribute to the higher grade.

6.5.4. RQ4: revealing the temporal dynamics between the phase transitions

To further explore SRL processes, we investigated the time dimension of SRL processes. In doing so, we calculated the gap between the SRL phases. We defined the gap as the interval between the last activity of a phase and the first activity of the next phase. In Figure 6-6, we charted the gap trend between phase transition, including D→P, P→E, and E→A, in the three clusters of CL2, CL3, and CL4. These three transitions represent one complete SRL cycle. Next, we examined the trend of A→E transition in CL3 and CL4 to identify a proper time for intervention to encourage students to engage in further study by applying their adapted strategies in further execution, i.e., to switch the cluster from CL2 to CL3. Finally, we examined the timing of the second planning phase that existed in CL4 (Figure 6-8), again, with the goal to identify after what elapsed time to encourage revisiting the planning phase in order to engage in higher SRL processes, as those in CL4.

The trend of three transitions that construct complete SRL processes is provided in Figure 6-6. The Figure represents the percentage of the cumulative distribution of the gap time (in hours) for each transition. It is important to note that the transitions shown are the first-time occurrences in each SRL process. It can be seen that three clusters showed similar patterns in the transitions, except for students in CL2 who showed they are transitioning from task definition to planning phase (D→P) almost one day behind the students in CL3 and CL4. Regarding P→E and E→A transitions, over 75% of students in all clusters completed their transitions in 50 hours (slightly over two days). At this stage, it is difficult to identify a threshold time to differentiate between the cluster trends for providing an intervention to encourage students to obtain new strategies that lead to higher metacognitive activities as manifested in higher clusters.

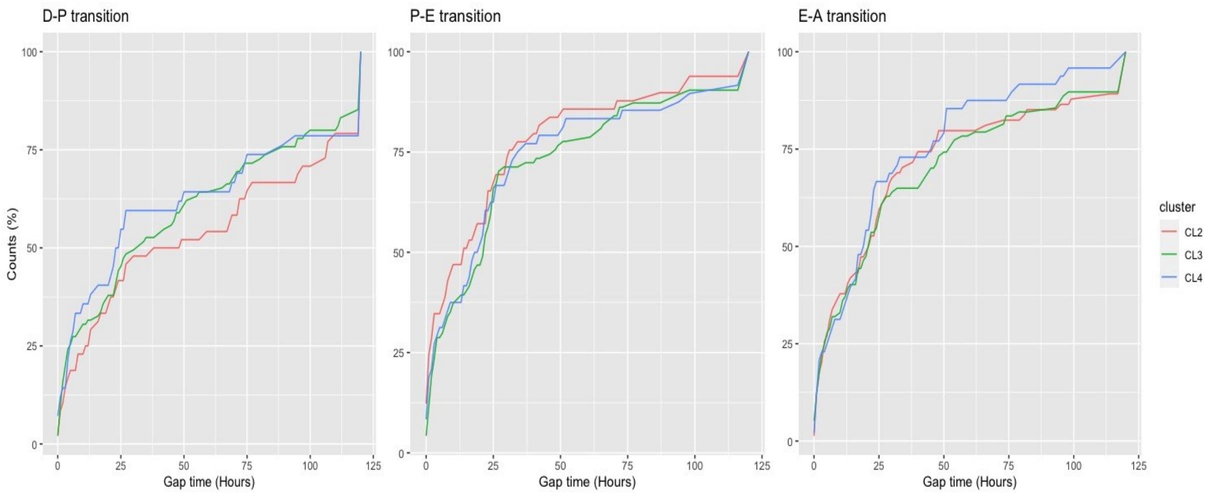

Figure 6-6. The trend in three transitions including D→P, P→E, and E→A with respect to their cluster.

However, in the transition of A→E, which only happened in CL3 and CL4, we can identify a clear trend (Figure 6-8). The trend suggests that, within 50 hours, 82.9% and 89.6% of students in CL3 and CL4, respectively, engaged in activities associated with the enactment phase *after* adapting phase. From the perspective of feedback, this timing can be used to encourage students who did not show the transition to consider engaging in the iterative execution phase after adapting their strategies.

The final time analysis looked at CL4, where SRL processes include the iteration in the planning phase. We measured the time interval between the first and second planning phases in this cluster. The finding suggested that, within 4 days, 50% of the students engaged in the second iteration in the planning phase and 85.4% within 8 days (Figure 6-8).

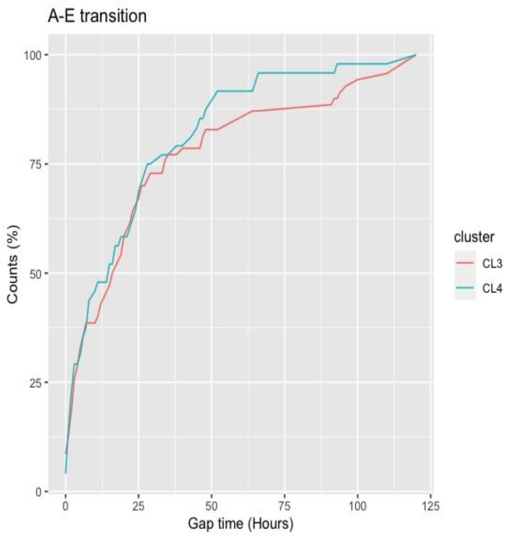

Figure 6-8 (a). The iteration of Adapting and Enactment phase.

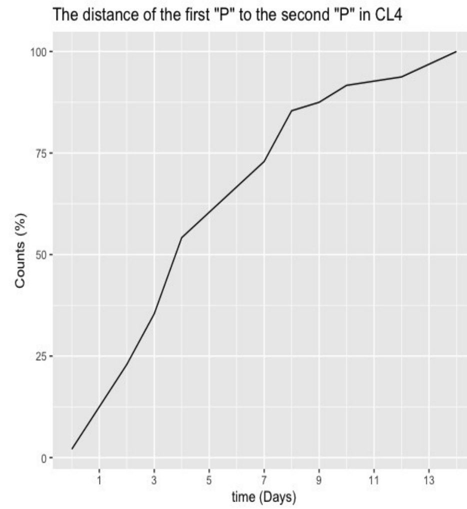

Figure 6-8 (b). The interval between the Planning phases.

6.6. Discussion

We studied students' SRL processes in the context of the programming course. In this study, we utilized the data representing SRL phases; the phases were detected from LMS log data using SRL indicators validated in our previous work (Salehian Kia et al., 2021). We identified four SRL process types using the clustering technique, which were relatively evenly distributed into four clusters. The clustering technique characterized the SRL processes based on the patterns of phase transitions and complexities in phases iterations, starting from incomplete processes (CL1) to processes with iteration in planning phases (CL4).

We triangulated the results from clustering with those from the classification determined by human coders who distinguished four types of SRL processes based on the theoretical-pragmatic meaning of the sequences (Hatala et al., 2023a) rather than on the computational grouping based on length and composition (i.e., what clustering algorithms do). The high agreement rate between the human coders and the clustering technique verified the affordance of the clustering method to characterize the processes that are structurally close to the theorized ideal SRL process, or deviations expected from the ideal process for less self-regulated learners. This result increases confidence in implementing automated systems for SRL process detection. Additionally, by triangulating the accuracy of detected processes, our research goes beyond other studies that relied on clustering techniques to discern learning strategies and tactics (Fan & Saint, 2021; Huang & Lajoie, 2021; Matcha, Gašević, Uzir, Jovanović, & Pardo, 2019). However, although some studies justified selecting SRL-related activities in the students' log data with theory, the lack of further elaboration from the SRL theory perspective in terms of processes students enact is noticeable in their findings (e.g. (Jovanović et al., 2017b; Moreno-Marcos et al., 2020)). Therefore, one of the contributions of this study is the elaboration of the SRL process patterns with respect to the SRL theory and the cyclical nature of SRL processes in particular.

The significance of discovered association between assignment grades with the SRL process clusters suggests the importance of enacted complexity of the phase iterations within SRL processes. While the importance of the quality of students' SRL processes on students' outcomes has been verified by the vast body of research (de Barba et al., 2020; Jovanović, Gašević, Pardo, Dawson, & Whitelock-Wainwright, 2019; Winne & Hadwin, 1998), our findings specifically highlight their cyclical nature. Students with incomplete SRL processes (CL1) seem to struggle with the materials and do not demonstrate basic SRL skills. On the other hand, the students in more advanced SRL process clusters showed more complex SRL behaviors, albeit to a different degree. This raises the question of to what extent the phase iteration in the SRL process can reveal higher learning gain. This issue demands more attention in educational studies, and future qualitative studies can reveal how students think when they enact SRL processes with high complexity in terms of phase iterations.

One of the major contributions of this study is stepping beyond exploring how SRL processes unfold and drawing attention to the timing aspect of SRL processes, which is critical in determining a proper time for an intervention. From this perspective, providing situated information feedback is more effective than providing passive information at the beginning (Hattie & Timperley, 2007). One of the aims of this study was to pave the way for enacting intervention by examining when, during a progressing SRL process as a student works on an assignment, we can determine that the student's process will not evolve into the more complex one. To do so, we examined SRL phase transition gap times at the process type level (clusters). We showed a general time trend for each transition; however, we did not find major time differences between clusters in terms of transitions from D→P, P→E, and E→A. However, we observed that most transitions from adapting to enactment (A→E), an indicator of the more advanced SRL clusters CL3 and CL4, occurred mainly within 48 hours after one complete SRL cycle. This can point to a possible intervention time prompting students towards strategies such as reviewing their products and comparing them with the standards (P. Winne & Hadwin, 1998), or other interventions more closely related to the learning task topic, i.e., programming strategies. Secondly, the highest level of SRL processes with the iteration of two planning phases happened within 8 days from the first planning phase. As this is a delay quite late in the 2-week assignment cycle, an intervention 8 days before the deadline (or earlier) to initiate the first planning cycle is indicated. Having this information, well-designed feedback either from an automated system via prompts, or instructor's intervention can lead the student to improve SRL skills. Our findings about the timing of interventions with respect to the concrete SRL activity go beyond the general strategies proposed in the literature (e.g., (Ifenthaler, 2012; Moos, 2017)) and demonstrate an approach that can be utilized to determine the appropriate time for personalized intervention based on a student's unraveling SRL process.

6.7. Conclusion

The framework plays a key role in shaping our study design, encompassing the identification of relevant studies and guiding the process of feature engineering. Thanks to the framework, we explored two aspects of temporality in students' SRL behavior to understand the dynamics of SRL phase transitions. The first aspect refers to the temporal order of SRL phases behaviors. We defined the SRL behaviors based on Winne and Hadwin's SRL theory which denoted four phases of SRL (P. Winne & Hadwin, 1998). We used indicators of SRL phases in the LMS clickstream data, validated in our previous study (Salehian Kia et al., 2021), to generate SRL processes, i.e., the sequence of phases. Using the clustering technique, this study discovered and characterized four types of SRL processes. The SRL types were clustered into the kinds of iterative behaviors over SRL phases which correspond to the theorized self-regulatory behaviors of students at different levels of SRL skills. We also found a significant association between SRL types and the assignment grades, suggesting the higher achieved learning outcomes, i.e., programming skills demonstrated in the assignments, being associated with more advanced SRL processes. Our result also revealed temporal dynamics between SRL phase transitions. Specifically, we showed that a two-day interval is an appropriate time to observe if the iteration of transition from adapting to enactment (A→E) happened and intervene if it was not observed. The transition is important because it manifests students' ability to regulate their learning behavior better; hence, its absence is an opportunity for an intervention.

Our study on the dynamics of temporal transitions in the SRL phases in a problem-solving context provides an opportunity for further research, especially for studies that focus on providing feedback and interventions. However, our approach needs to be validated in other types of tasks. Further research is needed to extend our approach to other types of data. This requires repeating design and considering the semantic meaning of log records with respect to SRL theory. One aspect that can be improved is an instructional design in LMS to improve the accuracy of data collected with respect to capturing the metacognitive and cognitive activities (Bernacki, 2018). Furthermore, we discussed how the identified SRL process types could explain differences between students. Further research can examine to what extent the difference in SRL phase cycles can represent a different degree of metacognitive activities and consequently lead to different learning outcomes.

Chapter 7. Case study 2: What Students See in the Dashboards and Learning Behavior That Follows

In this chapter, like the previous one, we aim to display and assess how the framework was employed while carrying out a temporal study. The main aim of the second case study is to investigate student behavior in response to feedback visualization. Specifically, we explore how the visualization impacts students' interactions with their subsequent postings. The insights from this study provide a deeper understanding of student' behaviors in online discussions which is important for personalizing communication. This personalization aims to determine the effective way to present information to students, thus improving the dashboards' ability to motivate positive behavioral changes.

7.1. Introduction

Student-facing dashboards are one of the main outcomes of research in learning analytics, aimed at informing students about their learning. Several systematic studies over the past few years have reviewed research and findings in LA dashboard research from various viewpoints (Bodily & Verbert, 2017; Matcha et al., 2020; Schwendimann et al., 2017). These studies have summarized key findings so far, identified main gaps in research, and offered recommendations for future research directions. These include, but are not limited to, the need to study perceived and actual effects on student learning and outcomes, execute randomized controlled trials or quasi-experimental studies, examine student perceptions and perceived effects on student behavior and achievement, and improve the quality of reporting on studies in several directions. The latest of these studies, by Matcha et al., proposed a model for a user-centered learning analytics system (MULAS) to guide and recommend researchers in establishing theory, designing dashboards, providing learners with feedback, and evaluating such systems (Matcha et al., 2020).

The challenges in this relatively young field are diverse, and a substantive body of empirical research is needed. This study aims to contribute to this effort with the shared goal of building a body of knowledge to provide personalized dashboards to students. Our research aligns with Matcha et al.'s (2020) recommendations, aiming to understand how dashboards influence student behavior, especially in online discussions. Unlike previous studies, we focus on an in-depth analysis of activity data beyond mere dashboard view counts.

7.2. Background

Having identified the scope of our work, we present the relevant background from existing research and state our guiding position in each of the reviewed areas.

7.2.1. Frame of Reference and Peer Comparison

In the context of the framework for pedagogical interventions including dashboards, Wise stressed the need to include frames of reference for students to understand their data in the dashboard (Wise, 2014). Jivet and colleagues identified three reference frames: 1) comparison with peers, 2) progress towards goals, and 3) own improvement (Jivet et al., 2017). Out of the three, comparison with peers has been the most prevalent frame used so far, as it is easy to implement and requires no additional input from the learner. In prior research, comparison with peers, as reviewed in Jivet et al., (2018), showed mixed preferences from learners. Although the results were not conclusive, the pattern that emerged was that high-achieving students benefited from comparison with their peers (Kim et al., 2016). This is contrasted with Corrin and Barba's results for class average students, where students above the class average became demotivated when seeing their status (Corrin & Barba, 2014). These results were based on using a talk-aloud protocol wherein students seeing the dashboards were asked ‘to articulate the actions that they would take in response’, rather than observing what students actually did after seeing the dashboard.

In this research, by analyzing dashboard content as seen by a student at a particular moment, we capture the student’s sense of success or failure. In our context of discussion activity, we posit that students do not consider their ability to contribute to the discussion to be different from their peers, as opposed to an activity that may require specialized disciplinary knowledge and skills, such as a programming assignment or creative writing, or when the dashboard shows the overall performance in the course, which often makes it harder for students to determine what they should do next.

This study addresses two research questions:

1. *RQ1: How does what students see in different states of dashboards influence their subsequent discussion-related learning activities?*

2. *RQ2: What are the associations between different states of dashboards and the timing patterns of students' re-engagement with discussion activities? (How different states in the dashboard might influence the timing of interaction with the discussion activities)*

7.3. The Use of the Framework to Conduct the Visualization Study

Similar to the previous chapter, our framework contributed to clarifying the steps to conduct a temporal study. We followed the information process proposed in the framework to reveal the temporality in posting behaviors in students after receiving visualization feedback. In this section, we demonstrate how the framework provided guidance for each step in conducting the study, as Figure 7-1 shows different components of the temporal studies and how to approach them. Since the data is already collected, we focus on the remaining two components: feature engineering and analysis and discussion.

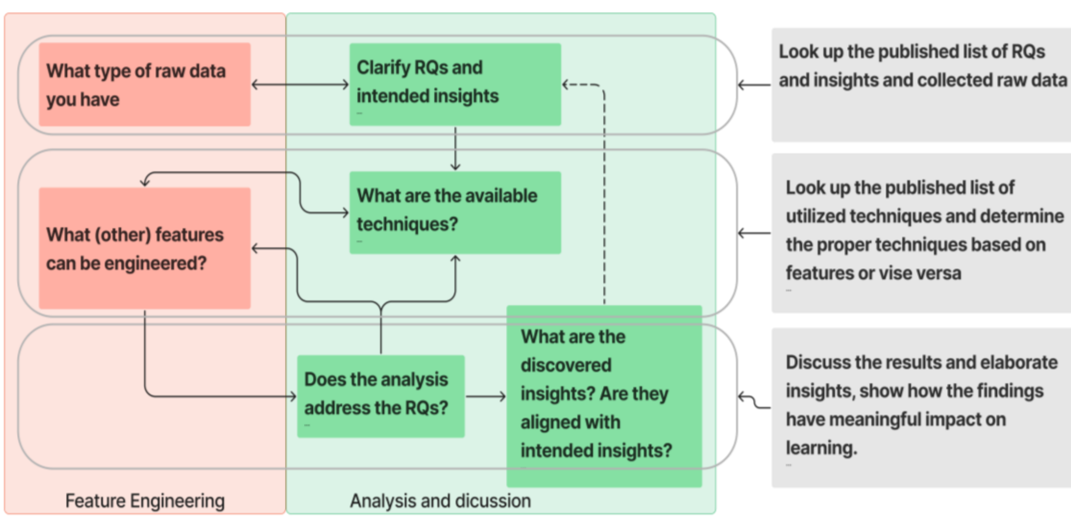

Figure 7-1. The framework process for the conducting temporal study.

7.3.1. Approaching RQs and Intended Insight Based on Available Data

The initial step is to clarify the aims of the study with respect to the availability of data. Specifically, we needed to assess whether the data at hand were sufficient to address our study's aim of offering insights into learning and, if so, the nature of these insights. This study utilized data previously collected to understand the impact of visualization feedback on students' subsequent activities, consisting of viewing the dashboard, posting in discussions, and viewing other students' posts. Therefore, we defined the main aim of this study as understanding the impact of dashboard feedback on students' subsequent posting behavior. A preliminary review of Table 4-11 suggests that the possible learning insight will reflect the impact of *feedback on learning* and *identify the learning indicator* that reveals students' engagement to assess the visualization's impact.

The *collaboration* insight is not the primary focus of this study because studies with the collaboration insight mainly explored the process of collaborative knowledge building toward reaching a goal. However, we did not completely exclude this category, as their methods for characterizing posting activities can be helpful for our study.

Next, we considered whether our existing data sufficiently supported this goal and examined how our research focus might align with the categories outlined in Table 4-5. Therefore, based on the data we have from our previous study on different types of visualizations (Beheshitha et al., 2016), it seems that our study RQs are aligned with the group of studies in the category *exploring comparing different groups of students* and possibly *time to intervention* and *exploring non-SRL learning indicators*, since we need to define and measure learning engagement among students.

7.3.2. Exploring the Available Techniques and Feature Engineering

With a clear direction for our study, the framework helped us identify the right techniques and refine our feature set. By using our web reference tool, we were able to filter out relevant studies based on attributes that matched ours. We first shortlisted studies focusing on insight about *learning indicator* or *feedback* (n=90), then further narrowed it down to those exploring *other than SRL OR group comparison* OR *exploring Scio-dynamics* (n=54), Then, we selected 44 papers that used LMS log data *OR* contextual data as their primary data source. Furthermore, it is apparent that our study contained *trace-forum* from tracing discussion student activities. This would help to further narrow down the result into 15 papers (Table below shows the result).

Table 7-1. List of relevant papers based on RQ focus, raw data, and aimed insight about learning.

ID	Title	Authors	Pattern level	Analytical technique
18	Towards Mutual Theory of Mind in Human-AI Interaction: How Language Reflects What Students Perceive About a Virtual Teaching Assistant	Wang, Qiaosi, Saha, Koustuv, Gregori, Eric, Joyner, David, Goel, Ashok	summative	basic.statistical.analysis
24	Supporting actionable intelligence: reframing the analysis of observed study strategies	Jovanovic, Jelena, Dawson, Shane, Joksimovic, Srecko, Siemens, George	event.sequence; group.event.pattern	process.mining; cluster.analysis ; visualization.analysis
35	Reply to which post? An analysis of peer reviews in a high school SPOC	Wang, Mengqian, Guo, Wenge, Le, Huixiao, Qiao, Bo	summative	basic.statistical.analysis
36	Learning Computational Thinking Without a Computer: How Computational Participation Happens in a Computational Thinking Board Game	Kuo, Wei Chen, Hsu, Ting Chia	transitional.pattern	process.mining; basic.statistical.analysis
76	Investigating students' interaction patterns and dynamic learning sentiments in online discussions	Huang, Chang-Qin, Han, Zhong-Mei, Li, Ming-Xi, Jong, Morris Siu-yung, Tsai, Chin-Chung	transitional.pattern	process.mining; visualization.analysis
82	How social challenges affect children’s regulation and assignment quality in hypermedia: a process mining study	Paans, Cindy, Onan, Erdem, Molenaar, Inge, Verhoeven, Ludo, Segers, Eliane	transitional.pattern	process.mining; visualization.analysis
103	A Mixed-Methods Approach to Analyze Shared Epistemic Agency in Jigsaw Instruction at Multiple Scales of Temporality	Oshima, Jun, Oshima, Ritsuko, Fujita, Wataru	other.sequential.patterns	network.analysis ; content.analysis; visualization.analysis
107	Understanding user behavioral patterns in open knowledge communities	Yang, Xianmin, Song, Shuqiang, Zhao, Xinshuo, Yu, Shengquan	transitional.pattern	process.mining

115	Effects of success v failure cases on learner-learner interaction	Tawfik, Andrew A, Giabbanelli, Philippe J, Hogan, Maureen, Msilu, Fortunata, Gill, Anila, York, Cindy S	transitional.pattern; summative	process.mining; content.analysis
123	Learner-generated materials in a flipped pronunciation class: A sequential explanatory mixed-methods study	Bakla, Arif	transitional.pattern; summative	process.mining
140	Promising Ideas for Collective Advancement of Communal Knowledge Using Temporal Analytics and Cluster Analysis	Lee, Alwyn Vwen Yen, Tan, Seng Chee	summative; group.event.pattern	content.analysis ; cluster.analysis ; visualization.analysis
150	A sequential analysis of responses in online debates to postings of students exhibiting high versus low grammar and spelling errors	Jeong, Allan, Li, Haiying, Pan, Andy Jiaren	transitional.pattern	process.mining
163	Impact of cultural diversity on students’ learning behavioral patterns in open and online courses: a lag sequential analysis approach	Tlili, Ahmed, Wang, Huanhuan, Gao, Bojun, Shi, Yihong, Zhiying, Nian, Looi, Chee Kit, Huang, Ronghuai	transitional.pattern	process.mining
166	Smart classroom environments affect teacher-student interaction: Evidence from a behavioural sequence analysis	Zhan, Zehui, Wu, Qianyi, Lin, Zhihua, Cai, Jiayi	transitional.pattern	process.mining
167	Putting It All Together: Combining Learning Analytics Methods and Data Sources to Understand Students’ Approaches to Learning Programming	Lopez,Äêpernas, Sonsoles, Saqr, Mohammed, Viberg, Olga	event.sequence; transitional.pattern	process.mining; frequent.sequence.mining; cluster.analysis

This stage helped us identify potential candidates for pattern-level feature engineering and select the appropriate techniques for our study. To further apply or derive inspiration from the information in the selected papers, we used guiding questions from chapter five, which assisted in refining our feature engineering and technique choices. For instance, from Table 7-1, we selected four papers closely aligned with our study (Huang et al., 2019; Jeong et al., 2017; Paans et al., 2019; Wang et al., 2020). We then explored the pattern-level features they employed, as detailed in Table 7-2. In considering features, we contemplated questions such as:

- *How will you define events based on the raw data?*
  Our primary features were derived from the LMS interaction clickstream, focusing on interaction with visualization pages and the discussion forum. Preparing this data required data cleaning and validation before conducting analysis (details are provided in the method section of this chapter).

- *Which specific traces (e.g., trace-exercise, trace-reading, trace-quiz) will be of primary interest?*
  We were particularly interested in specific types of student interactions, including those with visualization and posting. The activities are from trace-forum, which consisted of viewing visualization, viewing discussion, and posting discussion. By analyzing the sequences of these activities, we could infer aspects of the impact of visualization on subsequent posting.

- *Will you be extracting time features, and if so, what specific time-related attributes will you focus on (e.g., lag time, time taken)?*
  The study by (M. Wang et al., 2020) highlighted the importance of time in receiving attention from peers during posting and replying to discussion. We also considered time-related attributes, such as the duration and start time of each activity. This analysis allowed us to investigate the timing of viewing visualization and posting in discussion.

After studying works identified by our proposed framework (Table 7-2), techniques including process mining and visualization were most often used among the candidate studies. These techniques will be utilized for this case study.

Table 7-2. Candidates that are highly aligned with our study.

ID	Pattern-level Feature Engineering	Applicability to Our Study
35	Summative features: quantity of reviews to each post and post orders. Time of the review to the posts.	Time of posting can be a major contributor.
76	Transitional pattern: transitional probabilities of learning sentiments in different tasks.	Visualizing the transitional probabilities of different states.
82	Transitional pattern: transitions of the codes that represented the effects of social challenge in high social challenge dyads and using z-score to show the significance of the transitions.	The use of the Fuzzy Miner process mining algorithm, and z-score test.
150	Transitional pattern: transitional probability for the type and number of replies to each message type (with their corresponding z-scores).	Transitional probabilities (with z-score) could assess the posting response to visualization.

7.3.3. Analyzing Results and Refining Further RQs and Feature Engineering

Based on our chosen techniques and data, our findings for the visualization study highlighted various discussion behaviors among students. Our study has similarities to studies listed in Table 7-2. Addressing the guiding questions can help us assess and refine the findings:

- *How do your findings contribute to understanding the temporal aspects of learning?* This study contributes to a deeper understanding of the temporal patterns of behavior in discussions, specifically examining the effects of visualization feedback on the timing and nature of student discussion posts.

A notable aspect of our framework is its iterative nature, prompting us to continually evaluate and refine our results and insights. This iterative approach fosters innovation and deeper understanding. Addressing the following guiding questions (taken from Chapter 6) would help us in this respect.

- *What recommendations or interventions can be derived from the insights?*
  Our research indicates that visualizations have varied impacts on different students. It was observed that visualizations not only affect subsequent activities but also the timing of these activities in the context of discussion activities. Table 7-3 provides a summary of how the study was improved by considering this aspect. Furthermore, for future study, a more detailed examination of student-specific responses to visualizations is recommended. This approach would involve analyzing how individual differences among students affect their interaction based on individual motivation. Such studies could lead to personalized interventions, better suited to the unique learning profiles of each student, thereby enhancing both engagement and the efficacy of discussion-based learning activities.

- *Did the features that you utilized yield interesting results and insight that you aimed for? If not, do you need to revisit earlier stages for further feature extraction, or engineering?*
  In the first iteration, we uncovered patterns in activity transitions. However, we were also interested in timing (duration) aspects of those transitions. Therefore, more feature engineering is required to capture patterns in time intervals between and within transitions.

- *Are there any areas for improvement or further exploration based on the findings? Are there components in our framework that require revisiting?*

  To delve deeper into the time intervals of students' discussion activities, we should consider developing a new set of features that represent the timing of activities.

Table 7-3. Proposing improvement: Identifying Re-engagement Timings

What is Missing	New Features to Consider	Appropriate Technique
Understanding the timing of subsequent discussion	Calculating interval gap and transitions between activities.	Descriptive statistical analysis and visualization.

activity (re-engagement timing).		

In sum, our framework guided us in conducting the temporal research. It not only directed us in selecting appropriate analytical techniques and feature engineering methods but also prompted us to approach the study with deeper exploration in terms of re-engagmenet.

7.4. Method

7.4.1. Participants and Source of Data

Data for this study were derived from small-group discussion activities embedded in second- and third-year university courses. The courses, which included information technology, HCI, media production, game design, interactive arts, and media culture, all utilized blended delivery, employing an LMS to host course material and learning activities, including discussion forums. The discussions across these courses shared the same design, requirements, and course grade weighting. The dashboards, accessible from the top of the discussion page via a link or thumbnail, included text inviting students to check their progress.

Data were pooled from 63 discussion topics across four courses offered between Spring 2015 and Spring 2017. All students enrolled in the courses (n=107) participated

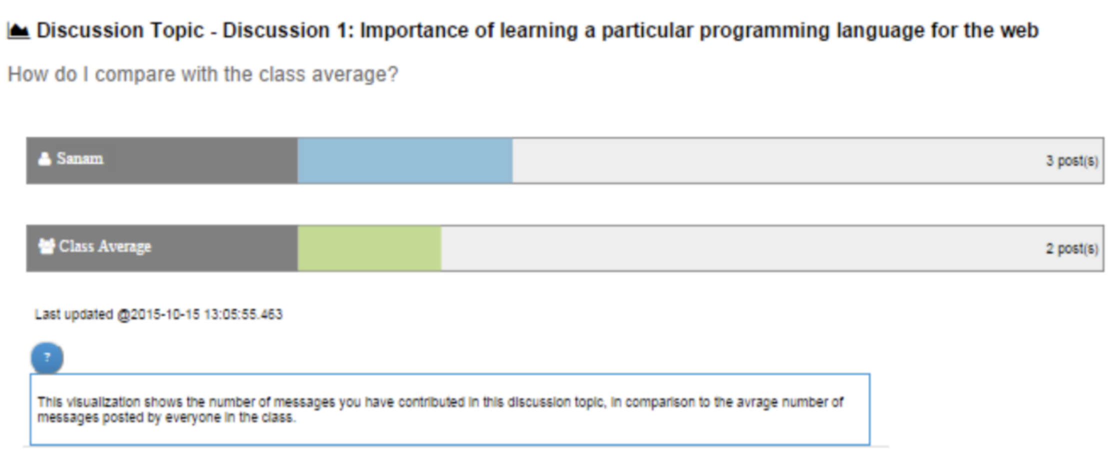

Figure 7-2. An snapshot of the visualization feedback.

in the discussions, resulting in 112 unique discussion/student participations, and their LMS log data were included in the analysis. A separate system for generating the dashboard (Figure 7-2), linked to the LMS via LTI protocol, logged all students' requests for viewing the dashboard and the view parameters.

7.4.2. Data Preprocessing and Encoding

The LMS displayed all discussion posts for a group on a single page. Viewing this page was recorded as a 'Read' action, representing reading the discussion. We coded both posting a top-level message (i.e., starting a new thread) and replying within a thread as a 'Post' action. Additionally, requests for displaying the dashboard were coded as 'View Dashboard'. To answer our RQs, we coded the 'View Dashboard' actions in a more detailed manner, depending on the information content seen by the student at that instance. Table 7-4 defines the codes used to categorize dashboard views.

Table 7-4. Students' actions and their definitions.

Action (Code label)		Description
View Discussion (D)		The student views the discussion page.
Post discussion (P)		The student posts a new discussion or replies to an existing discussion thread.
View dashboard Visualization	zero	The student views an empty dashboard, indicating not enough posts have been made in the class.
	below	The student sees their posting activity (the number of posts) is below the class average.
	at	The student sees their posting activity is at the class average.
	above	The student sees their posting activity is above the class average.

In the next steps, we created sessions of activities, defined by a 30-minute threshold of continuous discussion activities. We then coded these activity sessions based on the student’s actions within the session (Table 7-5 shows the coded activity session).

Table 7-5. Session of activities and their labels.

Activity Session (Code label)	Description
Only Discussion (D)	The session exclusively involves viewing discussions.
Posting (P)	The session includes at least one post and involves viewing discussions. Viewing discussion is included since students must open (view) the discussion first, then post it.
Viewing Visualization (zero, below, at, above)	The session involves viewing discussions and dashboard visualizations indicating the student's posting activity as Zero (no posts), Below, At, or Above the class average.
Posting + Viewing Visualization (Pzero, Pbelow, Pat, Pabove)	The session includes viewing discussions, making at least one post, and viewing dashboard visualizations at the state of Zero (no posts), Below, At, or Above the class average.

Having defined these sessions, Kim et al., (2016) suggested that analyzing the duration of these sessions and the gap intervals between them can serve as indicators of consistent engagement in discussion-based learning activities. Figure 7-3 depicts an example of the length and gap interval of the sessions of activities, where each block represents an activity session with its length and interval with the next session. In this example, the student initially views the discussion. During the subsequent session, the student visits the visualization, which displays no activity due to the absence of posts. In that same session, the student contributes at least one post, which is labeled as 'Pzero'. This label is used because the student does not revisit the updated visualization, and the last known status showed zero postings. In a subsequent session, the student revisits the visualization, which now displays postings below the class average. Since the student also posts in the discussion during this session, it is labeled 'Pbelow'.

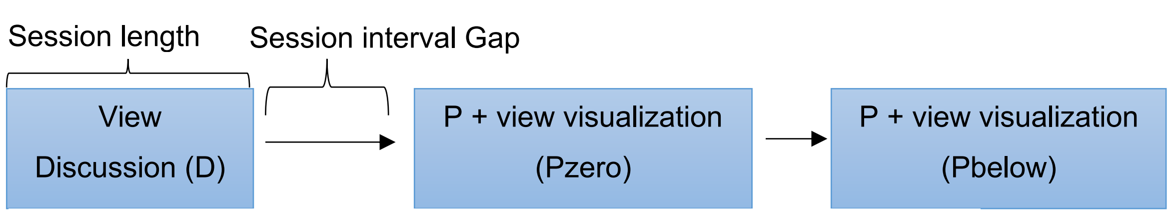

Figure 7-3. An example of sessions of activity for a student (case).

7.4.3. Data Analysis

We began with a descriptive data analysis to understand the general trends in our data and familiarize ourselves with it. Next, we conducted a lag sequential analysis (LSA) following Bakeman and Gottman's approach (Bakeman & Gottman, 1997). LSA is a statistical technique used for analyzing the order of sequential states (in our case, sessions), particularly in behavioral studies. It involves categorizing observed states and examining the likelihood of these states occurring in succession. The essence of LSA lies in identifying and assessing the significance of temporal patterns and dependencies between different state types (session types), providing insights into the dynamics of behavior sequences. The z-score is used to determine the significance of transitions between states, measuring how much the frequency deviates from the expected one. A high z-score indicates a significant, non-random relationship between states, revealing important behavioral patterns and dependencies in the sequence data. Additionally, examining higher-level lags (e.g., lag=2) allows us to explore how a state influences not the immediate next state but the one following it. This approach helps in understanding the longer-term influence of a state, uncovering indirect or delayed effects in the sequence of sessions, which can contribute to comprehensive behavioral analysis.

Prior to utilizing LSA, we met recommendations for minimum sample size (Bakeman & Gottman, 1997, p.114). To examine RQ1, the transitions between sessions (both lag=1 and lag=2) that occurred more frequently than expected with a z-score above 1.96 were visualized and interpreted. For RQ2, our approach involved a comparative analysis of session lengths and the intervals between sessions. This comparison was based on the type of session activities, providing insights into how different dashboards might influence the timing of students' interaction with the discussion activities.

7.5. Result

7.5.1. Exploring data

Initially, we explored the data distribution to gain insight into the general trend. Figure 7-4 shows two charts. First, the distribution of visualization view counts per student indicates a rapid decrease in frequency as the view count rises. The mode appears to be at one view per student, with very few students having more than six views. Second, in the "Distribution of Forum Post Counts per Case" chart, we observe that the frequency of students with two forum posts is the highest, with a notable drop-off in frequency as the post count per student increases. Students with more than three posts are noticeably less prevalent.

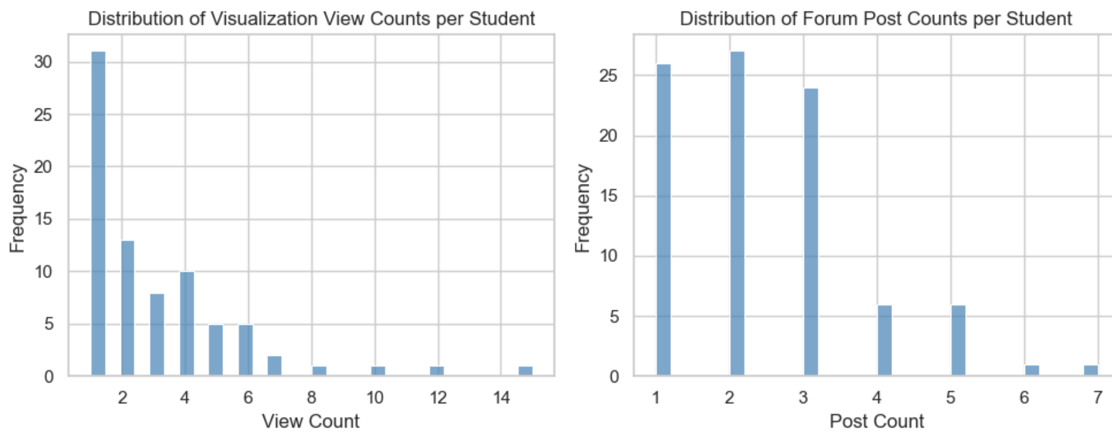

Figure 7-4. Distribution of view visualization and post counts (per student).

Next, we explored the distribution of session lengths (durations) and intervals between sessions. These sessions include discussion activities as outlined in the method section (Table 7-5). The average number of sessions per student was 14.0 (M=14.0, SD=9.8), with a median of 11.0 (Md=11.0). Figure 7-5 presents two distribution charts. The first chart shows session lengths, with the x-axis indicating the upper limit of each interval. For instance, '250 seconds' represents intervals from 0 to 250 seconds, and '500 seconds' from 250 to 500 seconds. This chart indicates a skew towards shorter sessions, peaking below 250 seconds (about 4 minutes), with a gradual decrease at 500 seconds (~8 minutes) and a rapid decline for longer sessions.

The second chart, 'Distribution of Gap Intervals (Hours),' illustrates the intervals between sessions which can represent re-engagement with the discussion activities. It shows a high frequency for shorter gaps, especially around one hour, and a decreasing frequency as the gap interval lengthens, indicating that students commonly have shorter intervals between sessions. We also set a cap of 120 hours for the gap interval. In general, this chart suggests that shorter intervals between sessions are more common among the students.

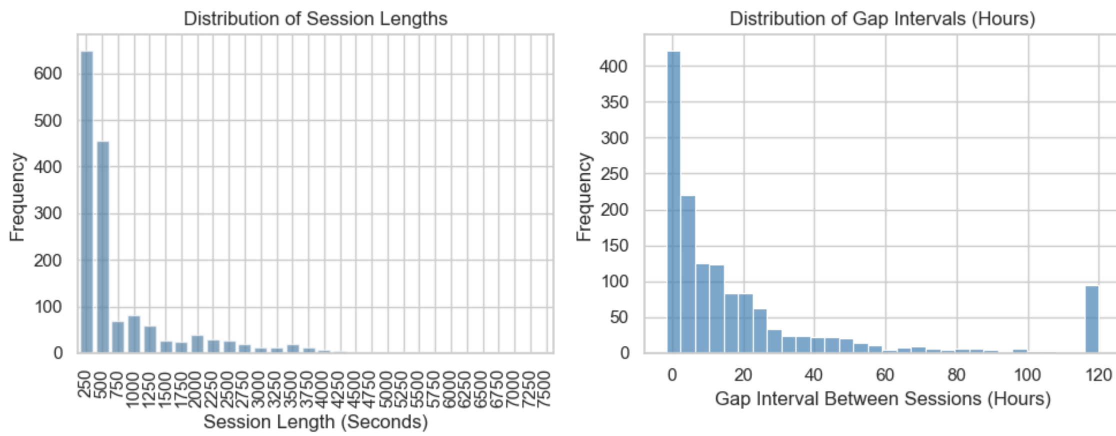

Figure 7-5. Distribution of session of activities' (e.g. Pbelow or Pat) lengths (durations), and interval between sessions.

Overall, these observations collectively provide insights into user engagement patterns, highlighting that shorter sessions and fewer posts per student are more prevalent. Additionally, students tend to revisit or engage with content within shorter intervals, typically around half an hour. In the next section, process analysis will differentiate how various types of visualizations may influence user behavior.

7.5.2. Process Mining

In this section, we address the first research question utilizing process mining. Figure 7-6 illustrates the outcomes of process mining, focusing on the z-score analysis of students' session transitions (lag=1). The y-axis represents the session type or start state, while the x-axis indicates the end state. Notably, sessions where students perceived themselves as being at the class average (labeled "at") were disproportionately followed by states where they 1) viewed the visualization again, and 2) continued to see themselves at the class average (column labeled "at", z = 12.67). Also, students who initially viewed their performance as below average (row labeled "below") demonstrated behaviors aimed at improvement, such as repeatedly engaging with visualizations and posting, depicting their progress from below to at the class average (columns, "P", z=3.41 and "Pat", z=3.72). This suggests that visualizations had a more encouraging effect on students who initially ranked below the average. Additionally, students who saw themselves above the average consistently monitored their relative standing, indicating an effort to maintain their superior position (row “Above” z=5.73, and “Pabove”, z= 3.83 to column ‘above’).

Figure 7-7 provides a heatmap visual representation of the z-scores for statistically significant two-state sequence transitions (lag=2). The y-axis represents the initial and subsequent state, while the x-axis indicates the resulting state. It is evident that students who initially saw themselves at the class average and then perceived a decline to below average (row labeled “at” → “below” or “Pbelow”) were more inclined to post to regain their average standing (column “Pat”, z= 11.5). Similarly, students frequently engaging in viewing discussions and then viewing themselves as below average (row “D” →“below”) were prompted to post (column “P”, z=3.7), aiming to

Figure 7-6. Depicts the transitions between sessions with a lag of 1. The y-axis represents the starting session, while the x-axis indicates the ending session. Displayed values are z-scores, where a score higher than 1.96 signifies a statistically significant transition, implying a frequency that is significantly higher than the expected value.

elevate their status back to the class average. The data also reveals a consistent pattern among students above the class average, where they took actions to maintain their position (row “above” → “Pabove” and “above” to column “above”, z=4.8, and 3.8, respectively). However, no significant transition was observed for students who dropped from above to at the class average, indicating these students may not have engaged with the visualizations. Regarding negative transitions, the data suggests that students who engaged in discussions and posting while viewing themselves as above class average were less likely to engage solely in posting during subsequent sessions.

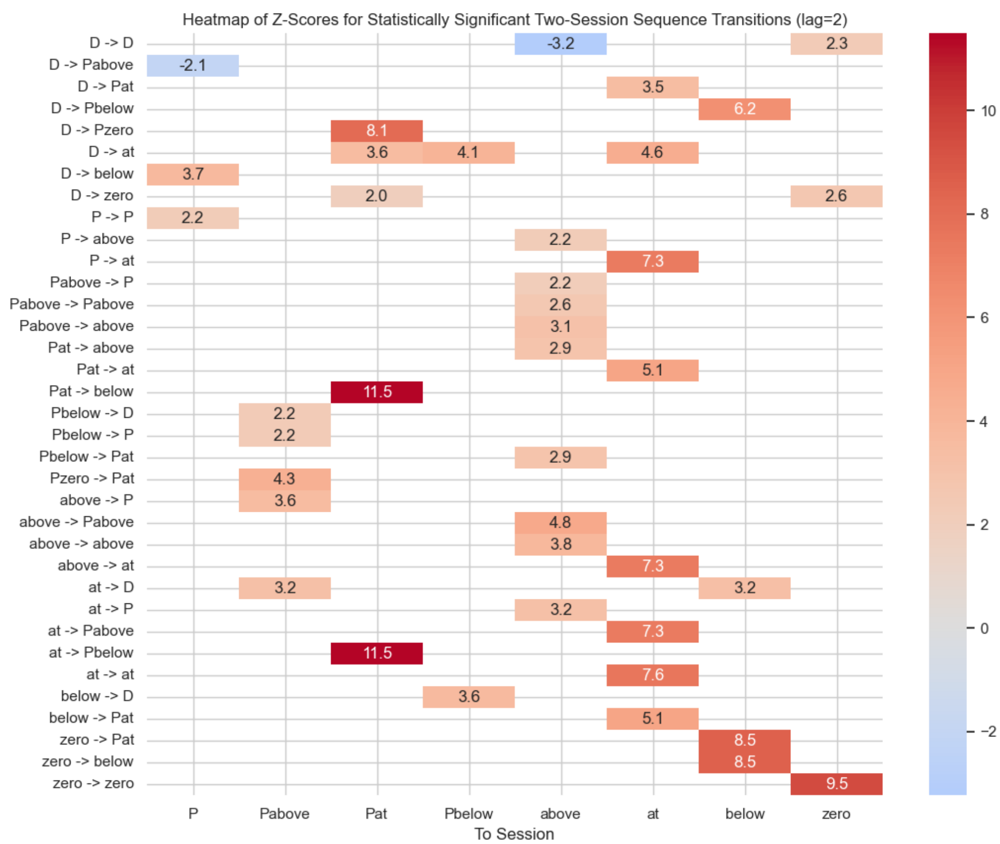

Figure 7-7. The transition heatmap which shows only significant transitions with lag=2. The y-axis shows the initial and the subsequent state, while the x-axis indicates the resulting state.

7.5.3. Timing analysis

Using timing analysis, we address our second research question. Figure 7-8 presents a comparison of session lengths across various engagement activities. Sessions involving viewing visualization without subsequent posting activities exhibit a median duration close to 20 minutes, with similar distribution ranges. In contrast, sessions that include posting activities generally span longer durations. Specifically, sessions combining posting with viewing visualizations categorized as Below average (labeled “Pbelow”) have the longest session lengths. The next longest sessions are those representing zero class average (labeled “Pzero”), indicating they were likely among the first posts in the discussion. Subsequent sessions include postings with visualizations at class average (“Pat”) and postings with visualizations above class average (“Pabove”). The data implies that students perceiving their performance as below the class average are inclined to spend more time in discussion forums, potentially to improve their understanding or performance.

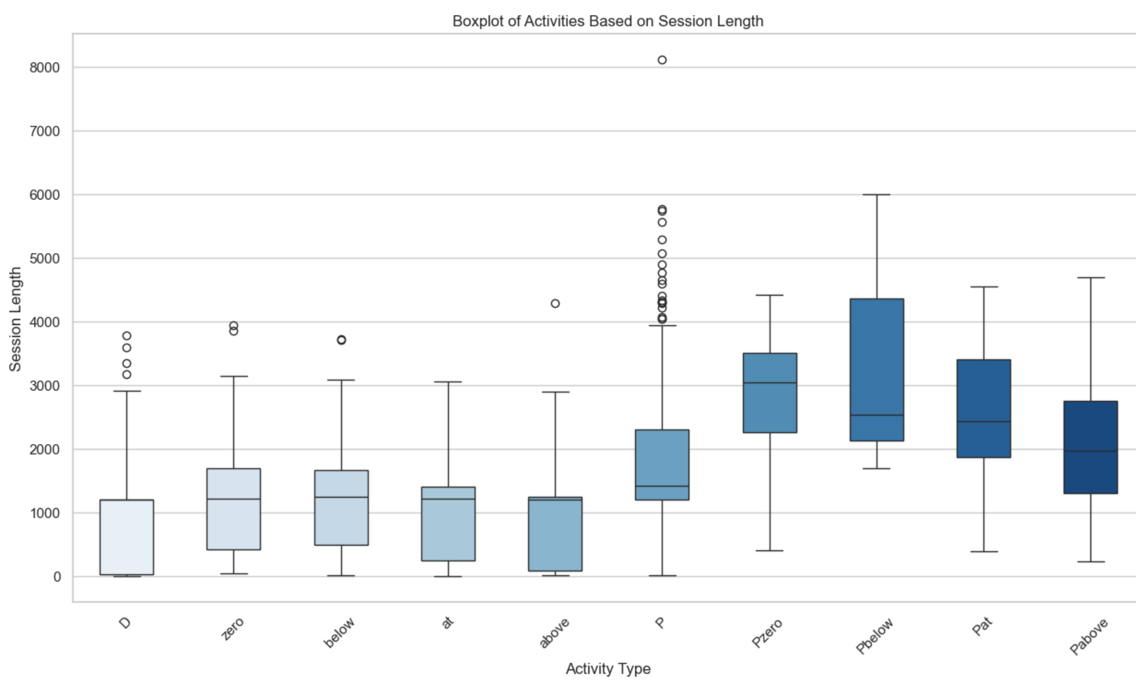

Figure 7-8. Session’s lengths (in seconds) based on their type.

Error! Reference source not found. presents the gap intervals after sessions categorized by session types. These intervals represent students’ re-engagement with the discussion activities. Notably, sessions involving posting and viewing visualization below the class average (labeled "Pbelow") were followed by the shortest intervals before subsequent sessions. This trend suggests a higher level of engagement among students in these sessions, with a propensity for quicker return to discussion. Conversely, sessions aligning with the class average for Visualization and Posting (denoted as "Pat") were followed by longer gaps before students re-engaged with the discussion. Regarding sessions designated as "Pabove", both the average gap interval and the overall distribution appeared similar to those sessions exclusively focused on Posting. Overall, this comparison provides insight into how different activity types may encourage students to re-engage with the discussion.

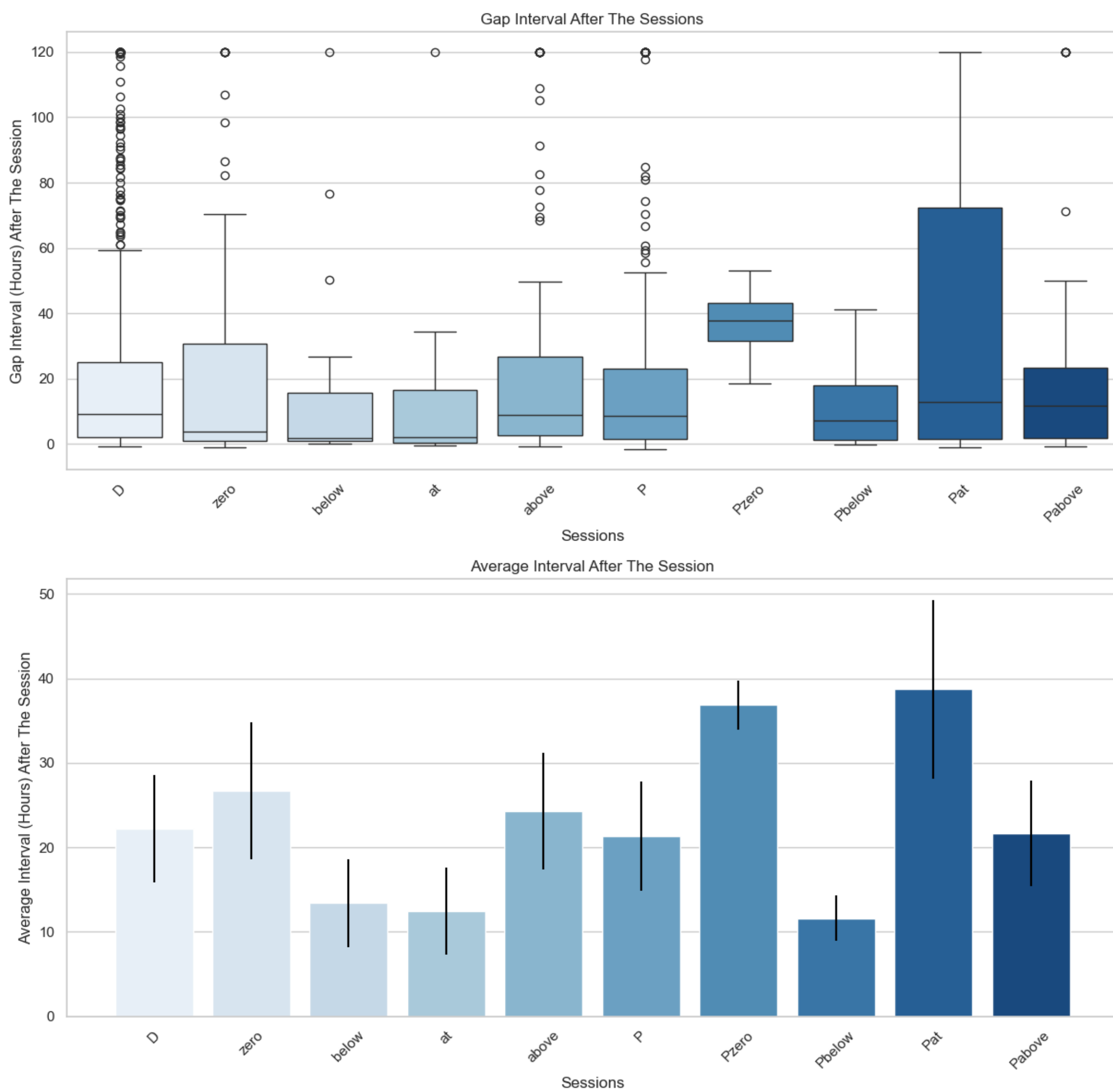

Figure 7-9. Gap intervals after each session which represent re-engagement of the students.

7.6. Discussion and Conclusion

In this section, we reflect on the findings introduced in the results sections. The main insight from this research is that dashboard users are likely to post more as a result of viewing the dashboard. We have discovered unique patterns of activity following dashboard views. These patterns differ if students see themselves at the lower, middle, or high end of the frame of reference used in the dashboard. We believe these patterns can form the basis for personalizing the process of designing dashboards for the user as the discussion unfolds and the user participates in the discussion.

7.6.1. Engagement Patterns Based on Self-Perception

Students "above" the class average tended to maintain their performance levels with frequent viewing but less content posting ("Above" → "Above", z=5.73). This behavior suggests a sense of satisfaction or a perceived lack of necessity to increase participation. In contrast, students who saw themselves as below average exhibited increased posting activity. This pattern indicates that the dashboard might have served as a motivational tool for these students, encouraging them to enhance their contributions to catch up with their peers. Additionally, students who initially considered themselves at the class average and subsequently observed a drop to below average actively endeavored to regain their former status (lag=2, transitions "At" → "PBelow" → "PAt", z=11.5). This shift demonstrates responsiveness to the feedback provided by the dashboard, further highlighting its impact on student engagement.

7.6.2. Temporal Dynamics of Student Engagement

The study also shed light on the temporal aspects of student engagement. Students who were below or at the average took longer to post after viewing the dashboard (longer session duration), suggesting they might be spending additional time in preparation and research (Figure 7-8). This deliberation could reflect a strategic approach to learning, where students invest more effort after realizing their relative standing. Also, these students took less time to return to discussion, showing that some states seen by the student have stronger motivating impact than the other states. (Figure 7-9). This quicker return indicates that the dashboard not only informs students of their standing but also acts as a catalyst for engagement, possibly due to increased awareness or a desire to improve their performance. Interestingly, students who initially posted more than average showed a subsequent decline in the length of their sessions, potentially indicating that initial high engagement without significant feedback or change in relative standing could lead to diminished effort over time.

7.6.3. Limitation and future direction

While this study provides valuable insights, there are some limitations to consider. The study's focus on frequency of discussion visits and posting count alone may not fully capture the quality or depth of student engagement. Future research could include qualitative aspects of posts or incorporate other metrics of engagement, such as the quality and complexity of contributions. Additionally, exploring the long-term effects of such feedback mechanisms on learning outcomes would be beneficial.

An additional consideration is the static nature of the dashboard's accessibility. students had to actively seek out the dashboard, which may not sufficiently engage those in need of more encouragement, particularly those below average. A more proactive approach could involve automatic pop-up notifications during subsequent discussion accesses, displaying the current state of discussion. This method might more effectively prompt engagement from students needing additional motivation.

However, this approach is theoretical for another reason. Although presenting students with their low standing compared to their peers seemed to have a motivational impact on improving their posting behavior, the question remains as to what extent the comparison affected their self-value (Gerber et al., 2018). This could potentially induce unnecessary anxiety. These aspects require further investigation using different methodologies. Also, we assumed that our task was simplified by the fact that students do not perceive their ability to contribute to the discussion (do research, formulate ideas, and build on ideas of others) as dramatically different from others. However, during designing dashboards for different types of activities, it's crucial to consider how learners perceive the difficulty of these tasks in relation to their own skills. Ensuring that the information on the dashboards is presented within an appropriate reference frame is key, as it enables learners to make constructive assessments about their abilities and progress.

Nevertheless, we have demonstrated that by studying the actual content students see in the dashboards, we can discover interesting behavioral patterns that are much more frequent than expected. Notably, students who viewed their performance as below or at the class average were motivated to increase their engagement in comparison to other dashboard states or when they were not viewing the dashboard. These findings highlight the important role of dashboard feedback in shaping student behavior and engagement. This study suggests that dashboards, when designed and utilized effectively, can serve as powerful tools for motivating students and enhancing their learning experience.

Chapter 8. Discussion on The Framework

Before finishing this thesis, I want to discuss our framework's development, how effective the framework was in guiding the research process in our two case studies, and its limitations. This chapter briefly covers how our framework was developed, compares it with current frameworks in learning analytics, and looks at its contributions and limitations as evaluated in the context of two case studies.

8.1. Reflection on the Framework Development

We developed our framework utilizing a data-driven approach, drawing from a substantial body of real-world research. It has similarities to the approaches used in the 'Method Framework for Design Science Research' (Johannesson & Perjons, 2021). In doing so, a systematic mapping study served as a foundational step in providing the data, and thus, insights needed to establish the framework. A systematic mapping study was essential in consolidating the current state of knowledge and providing a comprehensive overview of existing studies (Petersen et al., 2008). Our systematic mapping study identified different categories of research questions asked, data obtained, techniques utilized, and insights inferred about learning. It is noteworthy to mention that one key contribution of our framework was its focus on the granularity of the data, an aspect previously overlooked. The study also highlighted the connections between research questions, data, and analytical techniques, while considering the learning insights.

We adapted and extended Owen and Baker's (2020) framework to demonstrate how our mapping study's findings could aid researchers in data design, feature engineering, and analysis for implementing a temporal model. We followed Owen & Baker's structure, focusing on Data Design and Collection, Feature Engineering, Analysis, and Discussion. Based on our findings, we operationalized each component for temporal analytics studies in terms of questions asked, data used, applicable techniques, and insights gained, and developed a set of guiding questions for researchers to guide them through the process and a database of research works serving as an exemplar of various technique/data combinations.

Our framework has similarities and differences with other frameworks that have been recently proposed in LA. Saint and colleagues (2021) introduced the Trace-SRL framework, which transformed raw trace data into SRL events and then used process mining to capture SRL processes (Saint, Whitelock-Wainwright, et al., 2020b). Their framework is based on Siadaty et al., (2016) framework, which introduced a trace-based measurement protocol to measure the effects of scaffolding interventions on SRL processes. Using their framework, the study was able to pinpoint both effective and less effective SRL traits or strategies.

The structure of Saint's framework is centered around three primary components: (1) transforming raw trace data into SRL events, (2) selecting either supervised or unsupervised techniques for identifying learner types, and (3) applying process mining to investigate SRL processes. While our framework encompasses these components, Saint's framework is more narrowly focused on specific data and techniques utilized in SRL research. In contrast, our framework addresses a broader spectrum in temporal analytics. In terms of data transformation, Saint's framework is particularly detailed in its approach to data transformation for SRL, employing regular expression (REGEX) parsers to identify text patterns and define SRL event sequences. Conversely, our framework incorporates iterative feature engineering and facilitates the process with direct links to the exemplar literature, with a specific focus on granularity considerations in data transformation.

In terms of analytical techniques, Saint's framework elaborates how the choice between supervised and unsupervised techniques can reveal patterns in learners' SRL microprocesses. Our framework similarly guides the selection of analytical methods based on the data at hand, offering a distinct advantage in its broader applicability across various educational temporal studies.

Regarding process mining, Saint's framework exclusively utilizes the First-order Markov model (FOMM) to describe students' SRL strategies. In contrast, our framework not only incorporates extensive use of process mining as documented in the literature but also includes additional techniques. We placed a strong emphasis on complementary methods that can either corroborate process mining analyses or provide an additional layer of insight. This holistic approach to exploring various techniques is an aspect notably absent in Saint's framework.

In another framework, proposed by Hantoobi et al., (2021), the focus is on three main components: data collection, analysis, and application. They reviewed 19 papers and discussed five categories in their study. First, they stressed the need for diverse factors like platform interactions and feedback engagement to predict academic achievements. Second, they explored the link between theory and understanding learning. In other words, Hantoobi et al.'s framework is geared towards using predictive models to predict learning outcomes and applying learning theories to support these predictions. It provided a broad overview, mainly concentrating on predicting educational trends and understanding theoretical aspects of learning. In contrast, our framework specifically emphasized data engineering and the exploration of various analytical techniques. We offer a detailed guide for step-by-step processing of the temporal data in educational studies, paying close attention to data granularity and the application of diverse temporal techniques. This level of detail, particularly in handling and analyzing temporal data, was not provided by Hantoobi et al.'s framework.

The third category in Hantoobi et al.'s study referred to the importance of a clear framework for LA studies, which aids in curriculum design and understanding educational outcomes. While Hantoobi et al. concentrated on improving curriculum design and understanding educational outcomes, our framework, although not directly focused on design, guides researchers in identifying the types of data for analysis. This indirect contribution to data design helps in planning effective data collection. This approach is essential for researchers looking to deeply understand the complex temporal patterns in educational data, an area Hantoobi et al. did not extensively explore.

Finally, in the last two aspects, they emphasized the significance of frameworks for practicality in learning settings and the need for guidelines that consider data and take a holistic approach in applying learning analytics. These aspects were the focal points of our framework. Our contribution lies in establishing structured, holistic guidelines for conducting temporal studies, encompassing data collection, data engineering, analytical techniques, and deriving insights through an iterative process. Specifically, our framework provides tailored guidelines for researchers concentrating on temporal data analysis within educational contexts.

8.2. Application of the Framework in the Case Studies

It is essential to evaluate the proposed framework's usefulness and effectiveness through conducting a follow-up case study. This approach is a widely recognized and established method in scholarly literature (Greller & Drachsler, 2012; Saint, Whitelock-Wainwright, et al., 2020b). For instance, Saint and colleagues (2021) evaluated their framework's effectiveness through a follow-up SRL study, identifying both efficient and less effective SRL strategies among students. Similarly, we conducted two distinct temporal studies to evaluate the effectiveness of our framework.

Our framework has effectively clarified the necessary steps for conducting these temporal studies. We employed the information processing method suggested in the framework to analyze the temporality of student behaviors. The initial step was to define the aims of our study in relation to the available data. This clarity was essential in guiding the direction of our research, helping us categorize the raw data, research focus, and insights about learning. These categories were crucial for choosing the right techniques and refining feature sets, as detailed in sections 6.1 and 7.3 for each case study.

Furthermore, our web reference tool was useful in identifying relevant studies with attributes similar to ours. Both of our case studies used data from previously collected sources. With these papers in hand, the framework assisted in feature engineering and the selection of appropriate techniques. These papers served as a basis for our choices in feature engineering and analytical methods, supported by the framework's information process.

In the process of conducting case studies based on the guidelines of our framework, we observed a noticeable variation in the extent of relevant study coverage. In the SRL study, after selecting a subdomain within the categories of research focus, learning insights, and available data, numerous highly relevant SRL papers were identified (as noted in section 6.1). These papers, sharing similarities in study design and objectives, offered a diverse array of feature engineering options (e.g., creating sessions of SRL phases) and analytical techniques for examining SRL behaviors. Our framework and reference tool instilled confidence to conduct the study in a manner corroborated by the literature. The results we obtained resonated with the existing literature, and the framework facilitated reflection on these findings due to the high relevance of the selected papers.

In contrast, in the dashboard study, we found that there is a notable deficit in temporal analysis regarding the aftermath of dashboard interactions. The task of selecting appropriate subdomains for learning insights and research questions, in light of the available data, was inherently more abstract, often leading to overlapping subdomains. This complexity complicated the process of identifying relevant literature, resulting in a diverse range of papers. However, these papers predominantly utilized process mining, possibly reflecting the prominence of this technique in temporal analytics, as discussed in our mapping study (chapter 4). Given that process mining necessitates defining states and examining transitions between them, we encountered a lack of inspiring feature engineering approaches for state definition. Consequently, we relied on our own interpretations to logically define states. This challenge led us to extend our research beyond the initially provided references, where we discovered Kim et al. (2016)'s model, which uses session intervals as indicators of consistent engagement in discussion-based learning activities.

An important aspect of our framework is its iterative nature. This approach encouraged us to continuously evaluate and adjust our results and insights. We also formulated guiding questions at each stage to refine our feature engineering and technique selection. For example, in addressing the question, "What recommendations or interventions can be derived from the insights?", we noticed that the timing aspect of the SRL phases had not been thoroughly investigated, leading to a significant contribution in our SRL study where we identified the timing of SRL phases and opportunities for intervention (as explained in 6.1.3). Similarly, in the visualization dashboard study, we determined the optimal timing for re-engaging students in discussion activities.

Overall, our framework was beneficial in conducting our temporal research. It helped us in selecting suitable analytical techniques and feature engineering methods and helped us to incorporate an additional aspect of temporality into our studies.

8.3. Limitations of the Framework

In developing our framework, we conducted a meticulous process to gather and analyze research papers. However, it is essential to acknowledge that this process might not have captured every relevant study, for two reasons. First, within the course of this PhD research, we had resources to cover only the period of 2017-2022. During this time, certain techniques such as process mining were widely used, whereas others, like network analysis, were less frequently employed. Secondly, given the rapidly interdisciplinary nature of the LA field, new approaches are continually emerging. For instance, techniques like survivor analysis, borrowed from other disciplines (e.g. medical science and social science), have started making their mark in LA research (e.g., Chen, (2021)).

Our framework categorizes research questions (RQs), data types, analytical techniques, and insights based on a comprehensive mapping study. The mapping study revealed trends in the LA field toward certain data types and techniques. For example, a considerable number of papers in our reference database applied process mining to identify learning indicators. While this highlights well-established techniques, it also reflects a certain imbalance in the framework. Some techniques, rich in application examples, offer a wealth of information for researchers (e.g., process mining and frequent sequence analysis). On the other hand, there are emerging methodologies that are still in their early development stages and, consequently, offer limited examples. A case in point is the innovative use of Artificial Intelligence algorithms based on Hidden Markov Models (HMM), as explored in the work of Ouyang et al., 2023. This research is noteworthy for its examination of the adaptive and temporal characteristics of collaborative problem-solving. It sheds light on the multi-modal, dynamic, and synergistic aspects of group collaboration, offering insights into how an adaptive, self-organizing system emerges during the collaborative problem-solving process.

This imbalance is further compounded by the partly reliance of our framework on a reference database, which, if not regularly updated, risks becoming outdated, specifically in terms of categories and examples. The dynamic nature of LA research means that what we have captured is a snapshot in time. As the field grows, the database must evolve to include these new developments.

To address these limitations and ensure our framework remains comprehensive, we propose a strategy of continuous updating, particularly focusing on the less represented areas. This updating process should involve community-led investigations and collaborations, fostering a collective effort to integrate the most current research and diverse methodologies. By broadening the categories and enriching the framework with a wider array of techniques and applications, we aim to make it a more versatile and useful tool for the LA research community. This effort not only acknowledges the current limitations but also embraces the opportunity for growth, ensuring that our framework remains a valuable asset in the dynamic and interdisciplinary field of Learning Analytics.

Chapter 9. Conclusion

Temporal research on learning has recently received considerable attention. With the increasing focus on the temporal aspects of learning as an emerging area in education research, there is a new need for a guiding framework to help researchers navigate this area. Such a framework can provide clarity, mitigate redundancy, and offer a systematic way to approach and handle the challenges of temporal analytics.

This thesis makes several contributions to the expanding field of temporal educational analytics, and thus, to learning analytics. It encompasses two main contributions: a comprehensive mapping study and a systematically developed framework. Additionally, it offers two auxiliary contributions in the form of follow-up case studies.

One of the main contributions was the systematic mapping study that identified and categorized various elements of published research and investigated current trends in educational studies that specifically address the temporal dimension. Initially, this thesis provided a detailed review of prior mapping research and associated guidelines to ensure the validity of the mapping study. Subsequently, leveraging a thematic coding method, the study elaborated on patterns in temporal research components, including the research questions posed, the data collected at various granular levels, the analytical techniques used, and the insights about learning that were derived. The mapping study also highlighted the associations between the components.

Next, a systematic mapping study served as a foundational step in providing the data, and thus, the insights needed to establish the framework. Another main contribution of this thesis was the framework, which can be utilized by researchers in the process of data design, feature engineering, and analysis to implement temporal research. This thesis also provided a reference tool to guide researchers in selecting the proper analytical tools and feature engineering methods based on their research aims and data. Additionally, we proposed a set of guiding questions for each stage of using the framework to conduct a temporal study.

As a showcase for the applicability and evaluation of the framework, I conducted two temporal case studies. The first case study aimed to reveal the temporal nature of students’ SRL behaviors. I elucidated how the framework guidelines assisted in unveiling various temporal aspects related to SRL behaviors. Thanks to the framework, our SRL study discovered two facets of temporality. The first facet relates to the sequence of SRL phases as we discovered four categories of SRL processes based on phase transitions and the recurring nature of SRL. The next facet was related to the timing of SRL phase transitions, which paves the way for intervention.

Similarly, the second case study utilized the information process proposed in our framework to investigate temporality in students' discussion-related behaviors. The research community has recognized the importance of personalized feedback based on students’ needs. To achieve this goal, we need to deepen our understanding of students' behaviors in discussion that can be used for personalization, in order to know what information to communicate to the student, and how to frame and present it to make the dashboards more effective in motivating students and leading to desirable behavioral changes. In this study, I elucidated how our framework guided each step in conducting the study, leading to the discovery of the impact of visualization on the dynamics of engagement within online discussion activities. Indeed, this study leveraged diverse aspects of temporality to comprehend how visualization feedback can affect students' participation in discussion postings, especially for students who saw themselves as behind their peers.

In both case studies, our framework guided us to employ methods that offer a more enriched perspective of learners' temporal behaviors, extending beyond the commonly used correlational or cross-sectional approaches. The insights gained from these two case studies not only provide practical implications for designing more effective SRL feedback systems (first case study) or online discussion feedback dashboards (second case study) but also reinforce the value and adaptability of our proposed framework in different learning contexts.

One of the main features of this framework is its comprehensive approach to guiding the analysis of temporal data. Unlike existing models, it does not focus narrowly on a specific aspect of temporal analytics but offers a broader perspective. This inclusivity makes the framework versatile and applicable to a wide range of studies within the field.

This thesis also discussed the limitations and short comes of the framework in its application based on our case studies, highlighting areas where improvements or adjustments might be needed for future research.

In summary, this thesis advances the field of temporal educational analytics forward by mapping the current state of the field and providing a comprehensive framework for conducting temporal studies.

References

Akkermans, H., Wielinaa, B., & de Hoog, R. (1994). CommonKADS: A Comprehensive Methodology for KBS Development. *IEEE Expert-Intelligent Systems and Their Applications*, *9*(6), 28–37. https://doi.org/10.1109/64.363263

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

Azevedo, R., Taub, M., & Mudrick, N. V. (2017). Understanding and Reasoning about Real-Time Cognitive, Affective, and Metacognitive Processes to Foster Self-Regulation with Advanced Learning Technologies. *Handbook of Self-Regulation of Learning and Performance*, 254–270. https://doi.org/10.4324/9781315697048-17

Bakeman, R., & M.Gottman, J. (1997). Observing interaction. In *Event (London)*. https://doi.org/10.1017/CBO9780511527685

Baker, R. S. J. D., Corbett, A. T., & Wagner, A. Z. (2006). Human Classification of Low-Fidelity Replays of Student Actions. *Proceedings of the Educational Data Mining Workshop at the 8th International Conference on Intelligent Tutoring Systems*, 29–36. http://130.203.136.95/viewdoc/summary?doi=10.1.1.102.2435

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

Basit, T. N. (2010). Manual or electronic? The role of coding in qualitative data analysis. *Https://Doi.Org/10.1080/0013188032000133548*, *45*(2), 143–154. https://doi.org/10.1080/0013188032000133548

Beheshitha, S. S., Gašević, D., & Hatala, M. (2015). *A process mining approach to linking the study of aptitude and event facets of self-regulated learning*. 265–269. https://doi.org/10.1145/2723576.2723628

Beheshitha, S. S., Hatala, M., Gašević, D., & Joksimović, S. (2016). *The Role of Achievement Goal Orientations when Studying Effect of Learning Analytics Visualizations*. 54–63. https://doi.org/10.1145/2883851.2883904

Ben-Eliyahu, A., & Bernacki, M. L. (2015). Addressing complexities in self-regulated learning: a focus on contextual factors, contingencies, and dynamic relations. *Metacognition and Learning*, *10*(1), 1–13. https://doi.org/10.1007/S11409-015-9134-6/TABLES/1

Bernacki, M. L. (2018). Examining the Cyclical, Loosely Sequenced, and Contingent Features of Self-Regulated Learning : Trace Data and Their Analysis. *Handbook of Self-Regulation of Learning and Performance*, 370–387. https://doi.org/10.4324/9781315697048-24

Bienkowski, M., Feng, M., & Means, B. (2014). Enhancing teaching and learning through educational data mining and learning analytics: An issue brief. *Educational Improvement Through Data Mining and Analytics*, 1–60.

Bodily, R., & Verbert, K. (2017). Trends and issues in student-facing learning analytics reporting systems research. *Proceedings of the Seventh International Learning Analytics & Knowledge Conference on - LAK '17*, 309–318. https://doi.org/10.1145/3027385.3027403

Bogarín, A., Cerezo, R., & Romero, C. (2018). A survey on educational process mining. *Wiley Interdisciplinary Reviews: Data Mining and Knowledge Discovery*, *8*(1). https://doi.org/10.1002/widm.1230

Boroujeni, M. S., & Dillenbourg, P. (2019). Discovery and temporal analysis of MOOC study patterns. *Journal of Learning Analytics*, *6*(1), 16–33. https://doi.org/10.18608/jla.2019.61.2

Boroujeni, M. S., Hecking, T., Hoppe, H. U., & Dillenbourg, P. (2017). Dynamics of MOOC Discussion Forums. *Proceedings of the Seventh International Learning*

Analytics & Knowledge Conference, 128–137. https://doi.org/10.1145/3027385.3027391

Caprotti, O. (2017). Shapes of Educational Data in an Online Calculus Course. *Journal of Learning Analytics*, *4*(2), 76–90. https://doi.org/10.18608/jla.2017.42.8

Chen, B., Knight, S., & Wise, A. (2018a). Critical issues in designing and implementing temporal analytics. *Journal of Learning Analytics*, *5*(1), 1–9. https://doi.org/10.18608/jla.2018.51.1

Chen, B., Knight, S., & Wise, A. F. (2018b). Critical Issues in Designing and Implementing Temporal Analytics. *Journal of Learning Analytics*, *5*(1), 1–9. https://doi.org/10.18608/jla.2018.53.1

Chen, B., Resendes, M., Chai, C. S., & Hong, H. Y. (2017a). Two tales of time: uncovering the significance of sequential patterns among contribution types in knowledge-building discourse. *Interactive Learning Environments*, *25*(2), 162–175. https://doi.org/10.1080/10494820.2016.1276081

Chen, B., Resendes, M., Chai, C. S., & Hong, H. Y. (2017b). Two tales of time: uncovering the significance of sequential patterns among contribution types in knowledge-building discourse. *Interactive Learning Environments*, *25*(2), 162–175. https://doi.org/10.1080/10494820.2016.1276081

Chen, B., Resendes, M., Chai, C. S., & Hong, H. Y. (2017c). Two tales of time: uncovering the significance of sequential patterns among contribution types in knowledge-building discourse. *Interactive Learning Environments*, *25*(2), 162–175. https://doi.org/10.1080/10494820.2016.1276081

Chen, B., Wise, A. F., Knight, S., & Cheng, B. H. (2016a). Putting temporal analytics into practice: The 5th international workshop on temporality in learning data. *ACM International Conference Proceeding Series*, *25-29-Apri*, 488–489. https://doi.org/10.1145/2883851.2883865

Chen, B., Wise, A. F., Knight, S., & Cheng, B. H. (2016b). Putting temporal analytics into practice: The 5th international workshop on temporality in learning data. *ACM International Conference Proceeding Series*, *25-29-Apri*, 488–489. https://doi.org/10.1145/2883851.2883865

Chen, L. K. (2021). Timing of support in one-on-one math problem solving coaching: A survival analysis approach with multimodal data. *ACM International Conference Proceeding Series*, 553–558. https://doi.org/10.1145/3448139.3448197

Cheng, H. N. H., Liu, Z., Sun, J., Liu, S., & Yang, Z. (2017). Unfolding online learning behavioral patterns and their temporal changes of college students in SPOCs. *Interactive Learning Environments*, *25*(2), 176–188. https://doi.org/10.1080/10494820.2016.1276082

Chiou, G. L., Hsu, C. Y., & Tsai, M. J. (2019). Exploring how students interact with guidance in a physics simulation: evidence from eye-movement and log data analyses. *Interactive Learning Environments*. https://doi.org/10.1080/10494820.2019.1664596

Choi, Y., Lee, Y., Cho, J., Baek, J., Kim, B., Cha, Y., Shin, D., Bae, C., & Heo, J. (2020). Towards an Appropriate Query, Key, and Value Computation for Knowledge Tracing. *Proceedings of the Seventh ACM Conference on Learning @ Scale*, 341–344. https://doi.org/10.1145/3386527.3405945

Corrin, L., & Barba, P. De. (2014). *Exploring students' interpretation of feedback delivered through learning analytics dashboards*. 629–633. https://research.monash.edu/en/publications/exploring-students-interpretation-of-feedback-delivered-through-l

de Barba, P. G., Malekian, D., Oliveira, E. A., Bailey, J., Ryan, T., & Kennedy, G. (2020). The importance and meaning of session behaviour in a MOOC. *Computers & Education*, *146*, 103772. https://doi.org/https://doi.org/10.1016/j.compedu.2019.103772

Dickersin, K., Scherer, R., & Lefebvre, C. (1994). Systematic Reviews: Identifying relevant studies for systematic reviews. *BMJ*, *309*(6964), 1286. https://doi.org/10.1136/bmj.309.6964.1286

Ding, M., Wang, Y., Hemberg, E., & O'Reilly, U.-M. (2019). Transfer Learning Using Representation Learning in Massive Open Online Courses. *Proceedings of the 9th International Conference on Learning Analytics & Knowledge*, 145–154. https://doi.org/10.1145/3303772.3303794

Dominguez, C., Garcia-Izquierdo, F. J., Jaime, A., Perez, B., Rubio, A. L., & Zapata, M. A. (2021). Using Process Mining to Analyze Time Distribution of Self-Assessment and Formative Assessment Exercises on an Online Learning Tool. *IEEE Transactions on Learning Technologies*, *14*(5), 709–722. https://doi.org/10.1109/TLT.2021.3119224

Du, X., Zhang, M., Shelton, B. E., & Hung, J. L. (2019). Learning anytime, anywhere: a spatio-temporal analysis for online learning. *Interactive Learning Environments*. https://doi.org/10.1080/10494820.2019.1633546

Engerer, V. P. (2020). Implementing dynamicity in research designs for collaborative digital writing. *Education and Information Technologies*. https://doi.org/10.1007/s10639-020-10365-3

Falkner, K., Szabo, C., Vivian, R., & Falkner, N. (2015). Evolution of Software Development Strategies. *Proceedings - International Conference on Software Engineering*, *2*, 243–252. https://doi.org/10.1109/ICSE.2015.153

Falkner, K., Vivian, R., & Falkner, N. J. G. G. (2014). Identifying Computer Science Self-Regulated Learning Strategies. *Proceedings of the 2014 Conference on Innovation & Technology in Computer Science Education - ITiCSE '14*, 291–296. https://doi.org/10.1145/2591708

Fan, Y., & Saint, J. (2021a). A learning analytic approach to unveiling self-regulatory processes in learning tactics. *Journal of Learning Analytics*, 184–195.

Fan, Y., & Saint, J. (2021b). A learning analytic approach to unveiling self-regulatory processes in learning tactics. *Journal of Learning Analytics*, 184–195.

Fan, Y., Saint, J., Singh, S., Jovanovic, J., & Gašević, D. (2021a). A Learning Analytic Approach to Unveiling Self-Regulatory Processes in Learning Tactics. *LAK21: 11th International Learning Analytics and Knowledge Conference*, 184–195. https://doi.org/10.1145/3448139.3448211

Fan, Y., Saint, J., Singh, S., Jovanovic, J., & Gašević, D. (2021b). A Learning Analytic Approach to Unveiling Self-Regulatory Processes in Learning Tactics. *LAK21: 11th International Learning Analytics and Knowledge Conference*, 184–195. https://doi.org/10.1145/3448139.3448211

Fatahi, S., Shabanali-Fami, F., & Moradi, H. (2018). An empirical study of using sequential behavior pattern mining approach to predict learning styles. *Education and Information Technologies*, *23*(4), 1427–1445. https://doi.org/10.1007/s10639-017-9667-1

Fitz Gibbon, A., Joiner, D. A., Neeman, H., Peck, C., & Thompson, S. (2010). Teaching High Performance Computing to Undergraduate Faculty and Undergraduate Students. *Proceedings of the 2010 TeraGrid Conference*. https://doi.org/10.1145/1838574.1838581

Gabadinho, A., Ritschard, G., Müller, N. S., & Studer, M. (2011). Analyzing and Visualizing State Sequences in R with TraMineR. *Journal of Statistical Software*, *40*(4), 1–37. https://doi.org/10.18637/jss.v040.i04

Gaševic, 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

Gašević, D., Kovanović, V., & Joksimović, S. (2017). Piecing the learning analytics puzzle: a consolidated model of a field of research and practice. *Learning: Research and Practice*, *3*(1), 63–78. https://doi.org/10.1080/23735082.2017.1286142

Gerber, J. P., Wheeler, L., & Suls, J. (2018). A social comparison theory meta-analysis 60+ years on. *Psychological Bulletin*, *144*(2), 177–197. https://doi.org/10.1037/bul0000127

Greller, W., & Drachsler, H. (2012). Translating Learning into Numbers: A Generic Framework for Learning Analytics. In *Educational Technology & Society* (Vol. 15, Issue 3). http://groups.google.com/group/learninganalytics

Hansen, C., Hansen, C., Hjuler, N., Alstrup, S., & Lioma, C. (2017). Sequence modelling for analysing student interaction with educational systems. *Proceedings of the 10th International Conference on Educational Data Mining, EDM 2017*, 232–237.

Hantoobi, S., Wahdan, A., Al-Emran, M., & Shaalan, K. (2021). A review of learning analytics studies. In *Studies in Systems, Decision and Control* (Vol. 335, pp. 119–134). Springer Science and Business Media Deutschland GmbH. https://doi.org/10.1007/978-3-030-64987-6_8

Hatala, M., Nazeri, S., & Salehian Kia, F. (2023a). Progression of students' SRL processes in subsequent programming problem-solving tasks and its association with tasks outcomes. *The Internet and Higher Education*, *56*(March 2022), 100881. https://doi.org/10.1016/j.iheduc.2022.100881

Hatala, M., Nazeri, S., & Salehian Kia, F. (2023b). Progression of students' SRL processes in subsequent programming problem-solving tasks and its association with tasks outcomes. *The Internet and Higher Education*, *56*, 100881. https://doi.org/10.1016/J.IHEDUC.2022.100881

Hattie, J., & Timperley, H. (2007). The power of feedback. *Review of Educational Research*, *77*(1), 81–112. https://doi.org/10.3102/003465430298487/ASSET/IMAGES/LARGE/10.3102_00346543077001081-FIG1.JPEG

Hertz, M. (2010). What Do "CS1" and "CS2" Mean? Investigating Differences in the Early Courses. *Proceedings of the 41st ACM Technical Symposium on Computer Science Education*, 199–203. https://doi.org/10.1145/1734263.1734335

Hu, Q., & Rangwala, H. (2019). Reliable Deep Grade Prediction with Uncertainty Estimation. *Proceedings of the 9th International Conference on Learning*

Analytics & Knowledge - LAK19, 76–85. https://doi.org/10.1145/3303772.3303802

Hu, Y. H., Lo, C. L., & Shih, S. P. (2014). Developing early warning systems to predict students' online learning performance. *Computers in Human Behavior*, *36*, 469–478. https://doi.org/10.1016/j.chb.2014.04.002

Huang, C.-Q., Han, Z.-M., Li, M.-X., Jong, M. S., & Tsai, C.-C. (2019). Investigating students' interaction patterns and dynamic learning sentiments in online discussions. *Computers & Education*, *140*, 103589. https://doi.org/https://doi.org/10.1016/j.compedu.2019.05.015

Huang, L., & Lajoie, S. P. (2021a). Process analysis of teachers' self-regulated learning patterns in technological pedagogical content knowledge development. *Computers & Education*, *166*, 104169. https://doi.org/https://doi.org/10.1016/j.compedu.2021.104169

Huang, L., & Lajoie, S. P. (2021b). Process analysis of teachers' self-regulated learning patterns in technological pedagogical content knowledge development. *Computers & Education*, *166*, 104169. https://doi.org/https://doi.org/10.1016/j.compedu.2021.104169

Ifenthaler, D. (2012). Determining the effectiveness of prompts for self-regulated learning in problem-solving scenarios. *Educational Technology and Society*, *15*(1), 38–52. https://www.jstor.org/stable/jeductechsoci.15.1.38

Jeong, A., Li, H., & Pan, A. J. (2017). A sequential analysis of responses in online debates to postings of students exhibiting high versus low grammar and spelling errors. *Educational Technology Research and Development*, *65*(5), 1175–1194. http://www.jstor.org/stable/45018721

Jin, L., & Yu, D. (2019). Characteristics of Visual Attention for the Assessment of Conceptual Change: An Eye-Tracking Study. *Proceedings of the 10th International Conference on E-Education, E-Business, E-Management and E-Learning*, 158–162. https://doi.org/10.1145/3306500.3306584

Jivet, I., Scheffel, M., Drachsler, H., & Specht, M. (2017). Awareness is not enough. Pitfalls of learning analytics dashboards in the educational practice. *Lecture Notes in Computer Science (Including Subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics)*, *10474 LNCS*, 82–96. https://doi.org/10.1007/978-3-319-66610-5_7

Jivet, I., Scheffel, M., Specht, M., & Drachsler, H. (2018). License to evaluate: Preparing learning analytics dashboards for educational practice. *ACM International Conference Proceeding Series*, 31–40. https://doi.org/10.1145/3170358.3170421

Jo, I.-H., Kim, D., & Yoon, M. (2014). Analyzing the Log Patterns of Adult Learners in LMS Using Learning Analytics. *Proceedings of the Fourth International Conference on Learning Analytics And Knowledge*, 183–187. https://doi.org/10.1145/2567574.2567616

Johannesson, P., & Perjons, E. (2021). A Method Framework for Design Science Research. *An Introduction to Design Science*, 77–93. https://doi.org/10.1007/978-3-030-78132-3_4

Jovanović, J., Dawson, S., Joksimović, S., & Siemens, G. (2020). Supporting Actionable Intelligence: Reframing the Analysis of Observed Study Strategies. *Proceedings of the Tenth International Conference on Learning Analytics & Knowledge*, 161–170. https://doi.org/10.1145/3375462.3375474

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

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

Jovanović, J., Gašević, D., Pardo, A., Dawson, S., & Whitelock-Wainwright, A. (2019). Introducing meaning to clicks: Towards traced-measures of self-efficacy and cognitive load. *Proceedings of the 9th International Conference* *on Learning Analytics & Knowledge - LAK19*, 511–520. https://doi.org/10.1145/3303772.3303782

Kia, F. S., Teasley, S. D., Hatala, M., Karabenick, S. A., & Kay, M. (2020). How Patterns of Students Dashboard Use Are Related to Their Achievement and Self-Regulatory Engagement. *Proceedings of the Tenth International Conference on Learning Analytics & Knowledge*, 340–349. https://doi.org/10.1145/3375462.3375472

Kim, D., Park, Y., Yoon, M., & Jo, I. H. (2016). Toward evidence-based learning analytics: Using proxy variables to improve asynchronous online discussion environments. *Internet and Higher Education*, *30*, 30–43. https://doi.org/10.1016/j.iheduc.2016.03.002

Kim, J., Jo, I. H., & Park, Y. (2016). Effects of learning analytics dashboard: analyzing the relations among dashboard utilization, satisfaction, and learning achievement. *Asia Pacific Education Review*, *17*(1), 13–24. https://doi.org/10.1007/S12564-015-9403-8

Kinnebrew, J. S., Segedy, J. R., & Biswas, G. (2014). Analyzing the temporal evolution of students' behaviors in open-ended learning environments. *Metacognition and Learning*, *9*(2), 187–215. https://doi.org/10.1007/s11409-014-9112-4

Kitchenham, B. A., Budgen, D., & Pearl Brereton, O. (2011). Using mapping studies as the basis for further research - A participant-observer case study. *Information and Software Technology*, *53*(6), 638–651. https://doi.org/10.1016/j.infsof.2010.12.011

Knight, S., Friend Wise, A., & Chen, B. (2017). Time for Change: Why Learning Analytics Needs Temporal Analysis. *Journal of Learning Analytics*, *4*(3). https://doi.org/10.18608/jla.2017.43.2

Knight, S., Wise, A. F., & Chen., B. (2017). Time for change: Why Learning analytics needs temporal analysis. *Journal of Learning Analytics*, *4*(3), 7–17. https://doi.org/http://dx.doi.org/10.18608/jla.2017.43.2

Kokoç, M., Akçapınar, G., Hasnine, M. N., Kokoc, M., Akcapinar, G., & Hasnine, M. N. (2021). Unfolding Students' Online Assignment Submission Behavioral Patterns using Temporal Learning Analytics. *Educational Technology & Society*, *24*(1), 223–235. https://www.jstor.org/stable/26977869

Kuo, W. C., & Hsu, T. C. (2020). Learning Computational Thinking Without a Computer: How Computational Participation Happens in a Computational Thinking Board Game. *Asia-Pacific Education Researcher*, *29*(1), 67–83. https://doi.org/10.1007/s40299-019-00479-9

Lämsä, J., Hämäläinen, R., Koskinen, P., Viiri, J., & Mannonen, J. (2019). The potential of temporal analysis: Combining log data and lag sequential analysis to investigate temporal differences between scaffolded and non-scaffolded group inquiry-based learning processes. *Computers & Education*, 103674. https://doi.org/10.1016/J.COMPEDU.2019.103674

Lee, A. V. Y. (2021). Determining Quality and Distribution of Ideas in Online Classroom Talk using Learning Analytics and Machine Learning. *Educational Technology & Society*, *24*(1), 236–249. https://www.jstor.org/stable/26977870

Lee, A. V. Y., & Tan, S. C. (2017). Temporal Analytics with Discourse Analysis: Tracing Ideas and Impact on Communal Discourse. *Proceedings of the Seventh International Learning Analytics & Knowledge Conference*, 120–127. https://doi.org/10.1145/3027385.3027386

Liu, A. L. (2023). *Temporal, social, and dialogical characteristics of asynchronous online discussions and their implications for design* [Simon Fraser University]. https://summit.sfu.ca/item/36349

Liu, S., Kang, L., Liu, Z., Fang, J., Yang, Z., Sun, J., Wang, M., & Hu, M. (2021). Computer-supported collaborative concept mapping: the impact of students' perceptions of collaboration on their knowledge understanding and behavioral patterns. *Interactive Learning Environments*, 1–20. https://doi.org/10.1080/10494820.2021.1927115

Liu, Z., Yang, C., Rüdian, S., Liu, S., Zhao, L., & Wang, T. (2019). Temporal emotion-aspect modeling for discovering what students are concerned about in online course forums. *Interactive Learning Environments*, *27*(5–6), 598–627. https://doi.org/10.1080/10494820.2019.1610449

Loksa, D., & Ko, A. J. (2016). The Role of Self-Regulation in Programming Problem Solving Process and Success. *Proceedings of the 2016 ACM Conference on International Computing Education Research*, 83–91. https://doi.org/10.1145/2960310.2960334

Loksa, D., Ko, A. J., Jernigan, W., Oleson, A., Mendez, C. J., & Burnett, M. M. (2016). Programming, problem solving, and self-awareness: Effects of explicit guidance. *Conference on Human Factors in Computing Systems - Proceedings*, 1449–1461. https://doi.org/10.1145/2858036.2858252

Lum, P. Y., Singh, G., Lehman, A., Ishkanov, T., Vejdemo-Johansson, M., Alagappan, M., Carlsson, J., & Carlsson, G. (2013). Extracting insights from the shape of complex data using topology. *Scientific Reports*, *3*. https://doi.org/10.1038/srep01236

Lund, K., Quignard, M., & Williamson Shaffer, D. (2017). Gaining Insight by Transforming Between Temporal Representations of Human Interaction. *Journal of Learning Analytics*, *4*(3), 102–122. https://doi.org/10.18608/jla.2017.43.6

Lwande, C., Oboko, R., & Muchemi, L. (2021). Learner behavior prediction in a learning management system. *Education and Information Technologies*, *26*(3), 2743–2766. https://doi.org/10.1007/S10639-020-10370-6/TABLES/9

Mahzoon, M. J., Maher, M. Lou, Eltayeby, O., & Dou, W. (2018a). A Sequence Data Model for Analyzing Temporal Patterns of Student Data. *Journal of Learning Analytics*, *5*(1), 55–74. https://doi.org/10.18608/jla.2018.51.5

Mahzoon, M. J., Maher, M. Lou, Eltayeby, O., & Dou, W. (2018b). A Sequence Data Model for Analyzing Temporal Patterns of Student Data. *Journal of Learning Analytics*, *5*(1), 55–74. https://doi.org/10.18608/jla.2018.51.5

Mahzoon, M. J., Maher, M. Lou, Eltayeby, O., Dou, W., & Grace, K. (2018). A Sequence Data Model for Analyzing Temporal Patterns of Student Data.

Journal of Learning Analytics, *5*(1), 55–74. https://doi.org/10.18608/jla.2018.51.5

Matcha, W., Gašević, D., Ahmad Uzir, N., Jovanović, J., Pardo, A., Maldonado-Mahauad, J., & Pérez-Sanagustín, M. (2019). Detection of Learning Strategies: A Comparison of Process, Sequence and Network Analytic Approaches. *Lecture Notes in Computer Science (Including Subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics)*, *11722 LNCS*, 525–540. https://doi.org/10.1007/978-3-030-29736-7_39

Matcha, W., Gašević, D., Uzir, N. A., Jovanović, J., & Pardo, A. (2019a). Analytics of Learning Strategies: Associations with Academic Performance and Feedback. *Proceedings of the 9th International Conference on Learning Analytics & Knowledge*, 461–470. https://doi.org/10.1145/3303772.3303787

Matcha, W., Gašević, D., Uzir, N. A., Jovanović, J., & Pardo, A. (2019b). Analytics of Learning Strategies: Associations with Academic Performance and Feedback. *Proceedings of the 9th International Conference on Learning Analytics & Knowledge*, 461–470. https://doi.org/10.1145/3303772.3303787

Matcha, W., Uzir, N. A., Gasevic, D., & Pardo, A. (2020). A Systematic Review of Empirical Studies on Learning Analytics Dashboards: A Self-Regulated Learning Perspective. In *IEEE Transactions on Learning Technologies* (Vol. 13, Issue 2, pp. 226–245). Institute of Electrical and Electronics Engineers (IEEE). https://doi.org/10.1109/TLT.2019.2916802

Minović, M., & Milovanović, M. (2013). Real-Time Learning Analytics in Educational Games. *Proceedings of the First International Conference on Technological Ecosystem for Enhancing Multiculturality*, 245–251. https://doi.org/10.1145/2536536.2536574

Mohabbati, B., Asadi, M., Gašević, D., Hatala, M., & Müller, H. A. (2013). Combining service-orientation and software product line engineering: A systematic mapping study. *Information and Software Technology*, *55*(11), 1845–1859. https://doi.org/10.1016/j.infsof.2013.05.006

Molenaar, I. (2014a). *Advances in temporal analysis in learning and instruction*. *2*(4), 15–24. https://doi.org/10.14786/flr.v2i4.118

Molenaar, I. (2014b). *Advances in temporal analysis in learning and instruction*. *2*(4), 15–24. https://doi.org/10.14786/flr.v2i4.118

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

Moos, D. C. (2017). Emerging Classroom Technology : Using Self-Regulation Principles as a Guide for Effective Implementation. *Handbook of Self-Regulation of Learning and Performance*, 243–253. https://doi.org/10.4324/9781315697048-16

Moreno-Marcos, P. M., Muñoz-Merino, P. J., Maldonado-Mahauad, J., Pérez-Sanagustín, M., Alario-Hoyos, C., & Delgado Kloos, C. (2020). Temporal analysis for dropout prediction using self-regulated learning strategies in self-paced MOOCs. *Computers and Education*, *145*(May 2019). https://doi.org/10.1016/j.compedu.2019.103728

Morsy, S., & Karypis, G. (2019). Will this Course Increase or Decrease Your GPA? Towards Grade-aware Course Rec-ommendation. In *Journal of Educational Data Mining* (Vol. 11, Issue 2). https://doi.org/10.5281/ZENODO.3554677

Nazeri, S., Hatala, M., & Neustaedter, C. (2023). Associations of Research Questions, Analytical Techniques, and Learning Insight in Temporal Educational Research: A Systematic Mapping Study. *Journal of Learning Analytics*, *10*(2), 68–84. https://doi.org/10.18608/jla.2023.7745

Nazeri, S., Hatala, M., & Salehian Kia, F. (2023). When to Intervene? Utilizing Two Facets of Temporality in Students' SRL Processes in a Programming Course.

ACM International Conference Proceeding Series, 293–302. https://doi.org/10.1145/3576050.3576095

Neyem, A., Diaz-Mosquera, J., Munoz-Gama, J., & Navon, J. (2017). Understanding Student Interactions in Capstone Courses to Improve Learning Experiences. *Proceedings of the 2017 ACM SIGCSE Technical Symposium on Computer Science Education*, 423–428. https://doi.org/10.1145/3017680.3017716

Nistor, N., & Hernández-Garcíac, Á. (2018a). What types of data are used in learning analytics? An overview of six cases. *Computers in Human Behavior*, *89*, 335–338. https://doi.org/10.1016/J.CHB.2018.07.038

Nistor, N., & Hernández-Garcíac, Á. (2018b). What types of data are used in learning analytics? An overview of six cases. *Computers in Human Behavior*, *89*, 335–338. https://doi.org/10.1016/J.CHB.2018.07.038

Nowell, L. S., Norris, J. M., White, D. E., & Moules, N. J. (2017). Thematic Analysis: Striving to Meet the Trustworthiness Criteria. *Https://Doi.Org/10.1177/1609406917733847*, *16*(1). https://doi.org/10.1177/1609406917733847

Ochoa, X., Hershkovitz, A., Wise, A., & Knight, S. (2017). Towards a Convergent Development of Learning Analytics. *Journal of Learning Analytics*, *4*(3), 1–6. https://doi.org/10.18608/jla.2017.43.1

Ouyang, F., Xu, W., & Cukurova, M. (2023). An artificial intelligence-driven learning analytics method to examine the collaborative problem-solving process from the complex adaptive systems perspective. *International Journal of Computer-Supported Collaborative Learning*, *18*(1), 39–66. https://doi.org/10.1007/s11412-023-09387-z

Owen, V. E., & Baker, R. S. (2020). Fueling Prediction of Player Decisions: Foundations of Feature Engineering for Optimized Behavior Modeling in Serious Games. *Technology, Knowledge and Learning*, *25*(2), 225–250. https://doi.org/10.1007/s10758-018-9393-9

Paans, C., Onan, E., Molenaar, I., Verhoeven, L., & Segers, E. (2019). How social challenges affect children's regulation and assignment quality in hypermedia: a process mining study. *Metacognition and Learning*. https://doi.org/10.1007/s11409-019-09204-9

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

Pandey, S., & Srivastava, J. (2020). RKT: Relation-Aware Self-Attention for Knowledge Tracing. *Proceedings of the 29th ACM International Conference on Information & Knowledge Management*, 1205–1214. https://doi.org/10.1145/3340531.3411994

Petersen, K., Feldt, R., Mujtaba, S., & Mattsson, M. (2008, June 1). Systematic mapping studies in software engineering. *12th International Conference on Evaluation and Assessment in Software Engineering, EASE 2008*. https://doi.org/10.14236/ewic/ease2008.8

Petersen, K., Vakkalanka, S., & Kuzniarz, L. (2015). Guidelines for conducting systematic mapping studies in software engineering: An update. *Information and Software Technology*, *64*, 1–18. https://doi.org/10.1016/j.infsof.2015.03.007

Poitras, E. G., Doleck, T., Huang, L., Dias, L., & Lajoie, S. P. (2021). Time-driven modeling of student self-regulated learning in network-based tutors. *Interactive Learning Environments*, 1–22. https://doi.org/10.1080/10494820.2021.1891941

Qiao, C., & Hu, X. (2020). A joint neural network model for combining heterogeneous user data sources: An example of at-risk student prediction. *Journal of the Association for Information Science and Technology*, *71*(10), 1192–1204.

Reilly, J. M., & Dede, C. (2019). Differences in Student Trajectories via Filtered Time Series Analysis in an Immersive Virtual World. *Proceedings of the 9th*

International Conference on Learning Analytics & Knowledge, 130–134. https://doi.org/10.1145/3303772.3303832

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

Reimann, P., Markauskaite, L., & Bannert, M. (2014). E-Research and learning theory: What do sequence and process mining methods contribute? *British Journal of Educational Technology*, *45*(3), 528–540. https://doi.org/10.1111/bjet.12146

Riel, J., Lawless, K. A., & Brown, S. W. (2018a). Timing Matters: Approaches for Measuring and Visualizing Behaviours of Timing and Spacing of Work in Self-Paced Online Teacher Professional Development Courses. *Journal of Learning Analytics*, *5*(1), 25–40. https://doi.org/10.18608/jla.2018.51.3

Riel, J., Lawless, K. A., & Brown, S. W. (2018b). Timing Matters: Approaches for Measuring and Visualizing Behaviours of Timing and Spacing of Work in Self-Paced Online Teacher Professional Development Courses. *Journal of Learning Analytics*, *5*(1), 25–40. https://doi.org/10.18608/jla.2018.51.3

Roll, I., & Winne, P. H. (2015). Understanding, evaluating, and supporting self-regulated learning using learning analytics. *Journal of Learning Analytics*, *2*(1), 7–12. https://doi.org/10.18608/jla.2015.21.2

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. *ACM International Conference Proceeding Series*, 333–343. https://doi.org/10.1145/3448139.3448171

Saint, J., Gašević, D., Matcha, W., Uzir, N. A., & Pardo, A. (2020). Combining Analytic Methods to Unlock Sequential and Temporal Patterns of Self-Regulated Learning. *Proceedings of the Tenth International Conference on Learning Analytics & Knowledge*, 402–411. https://doi.org/10.1145/3375462.3375487

Saint, J., Whitelock-Wainwright, A., Gasevic, D., & Pardo, A. (2020a). Trace-SRL: A Framework for Analysis of Microlevel Processes of Self-Regulated Learning From Trace Data. *IEEE Transactions on Learning Technologies*, *13*(4), 861–877.

Saint, J., Whitelock-Wainwright, A., Gasevic, D., & Pardo, A. (2020b). Trace-SRL: A Framework for Analysis of Microlevel Processes of Self-Regulated Learning From Trace Data. *IEEE Transactions on Learning Technologies*, *13*(4), 861–877.

Salehian Kia, F. (2021). *Measuring Self-Regulatory Phases with Multi-Channel Trace Data in Open-Ended Learning Technology*. Simon Fraser University.

Salehian Kia, F., Hatala, M., Baker, R. S., & Teasley, S. D. (2021). Measuring students' self-regulatory phases in LMS with behavior and real-time self report. *ACM International Conference Proceeding Series*, 259–268. https://doi.org/10.1145/3448139.3448164

Saqr, M., & López-Pernas, S. (2021). The longitudinal trajectories of online engagement over a full program. *Computers and Education*, *175*, 104325. https://doi.org/10.1016/J.COMPEDU.2021.104325

Scherer, S., Weibel, N., Morency, L.-P., & Oviatt, S. (2012). Multimodal Prediction of Expertise and Leadership in Learning Groups. *Proceedings of the 1st International Workshop on Multimodal Learning Analytics*. https://doi.org/10.1145/2389268.2389269

Schwendimann, B. A., Rodríguez-Triana, M. J., Vozniuk, A., Prieto, L. P., Shirvani Boroujeni, M., Holzer, A., Gillet, D., & Dillenbourg, P. (2017). Understanding learning at a glance: A systematic literature review of learning dashboards. *IEEE Transactions on Learning Technologies*, *10*(1), 148–157. https://ieeexplore.ieee.org/stamp/stamp.jsp?tp=&arnumber=7542151

Sharma, K., Papamitsiou, Z., Olsen, J. K., & Giannakos, M. (2020). Predicting Learners' Effortful Behaviour in Adaptive Assessment Using Multimodal Data. *Proceedings of the Tenth International Conference on Learning Analytics & Knowledge*, 480–489. https://doi.org/10.1145/3375462.3375498

Sher, N., Kent, C., & Rafaeli, S. (2020). How “Networked” are Online Collaborative Concept-Maps? Introducing Metrics for Quantifying and Comparing the “Networkedness” of Collaboratively Constructed Content. *Education Sciences*, *10*(10), 1.

Sher, V., Hatala, M., & Gašević, D. (2019). On multi-device use: Using technological modality profiles to explain differences in students’ learning. *Proceedings of the 9th International Conference on Learning Analytics & Knowledge - LAK19*, 1–10. https://doi.org/10.1145/3303772.3303790

Shin, D., Shim, Y., Yu, H., Lee, S., Kim, B., & Choi, Y. (2021). SAINT+: Integrating Temporal Features for EdNet Correctness Prediction. *LAK21: 11th International Learning Analytics and Knowledge Conference*, 490–496. https://doi.org/10.1145/3448139.3448188

Siadaty, M., Gašević, D., & Hatala, M. (2016). Trace-Based Microanalytic Measurement of Self-Regulated Learning Processes. *Journal of Learning Analytics*, *3*(1), 183–214. https://doi.org/http://dx.doi.org/10.18608/jla.2016.31.11

Siemens, G., & Baker, R. S. J. D. (2012). Learning analytics and educational data mining: towards communication and collaboration. *Proceedings of the 2nd International Conference on Learning Analytics and Knowledge - LAK ’12*, 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

Studer, R., Benjamins, V. R., & Fensel, D. (1998). Knowledge engineering: Principles and methods. *Data & Knowledge Engineering*, *25*(1–2), 161–197. https://doi.org/10.1016/S0169-023X(97)00056-6

Sun, Z., Lin, C. H., Lv, K., & Song, J. (2021). Knowledge-construction behaviors in a mobile learning environment: a lag-sequential analysis of group differences.

Educational Technology Research and Development*. https://doi.org/10.1007/s11423-021-09938-x

Taub, M., & Azevedo, R. (2018). Using Sequence Mining to Analyze Metacognitive Monitoring and Scientific Inquiry based on Levels of Efficiency and Emotions during Game-Based Learning. In *Journal of Educational Data Mining* (Vol. 10, Issue 3). https://doi.org/10.5281/ZENODO.3554711

Theobald, M. (2021). Self-regulated learning training programs enhance university students' academic performance, self-regulated learning strategies, and motivation: A meta-analysis. *Contemporary Educational Psychology*, *66*, 101976. https://doi.org/10.1016/J.CEDPSYCH.2021.101976

Tsai, M.-J., Hou, H.-T., Lai, M.-L., Liu, W.-Y., & Yang, F.-Y. (2012). Visual attention for solving multiple-choice science problem: An eye-tracking analysis. *Computers & Education*, *58*(1), 375–385. https://doi.org/https://doi.org/10.1016/j.compedu.2011.07.012

Umer, R., Mathrani, A., Susnjak, T., & Lim, S. (2019). Mining Activity Log Data to Predict Student's Outcome in a Course. *Proceedings of the 2019 International Conference on Big Data and Education*, 52–58. https://doi.org/10.1145/3322134.3322140

Uzir, N. A., Gašević, D., Jovanović, J., Matcha, W., Lim, L.-A., & Fudge, A. (2020). Analytics of Time Management and Learning Strategies for Effective Online Learning in Blended Environments. *Proceedings of the Tenth International Conference on Learning Analytics & Knowledge*, 392–401. https://doi.org/10.1145/3375462.3375493

Van Goidsenhoven, S., Bogdanova, D., Deeva, G., Broucke, S. vanden, De Weerdt, J., & Snoeck, M. (2020). Predicting Student Success in a Blended Learning Environment. *Proceedings of the Tenth International Conference on Learning Analytics & Knowledge*, 17–25. https://doi.org/10.1145/3375462.3375494

Vieira, C., Parsons, P., & Byrd, V. (2018). Visual learning analytics of educational data: A systematic literature review and research agenda. *Computers & Education*, *122*, 119–135. https://doi.org/10.1016/J.COMPEDU.2018.03.018

Wang, M., Guo, W., Le, H., & Qiao, B. (2020). Reply to which post? An analysis of peer reviews in a high school SPOC. *Interactive Learning Environments*, *28*(5), 574–585. https://doi.org/10.1080/10494820.2019.1696840

Wang, Q., Saha, K., Gregori, E., Joyner, D., & Goel, A. (2021). Towards Mutual Theory of Mind in Human-AI Interaction: How Language Reflects What Students Perceive About a Virtual Teaching Assistant. *Proceedings of the 2021 CHI Conference on Human Factors in Computing Systems*. https://doi.org/10.1145/3411764.3445645

Wang, Y. Y., Li, T., Geng, C., & Wang, Y. Y. (2019). Recognizing patterns of student's modeling behaviour patterns via process mining. *Smart Learning Environments*, *6*(1). https://doi.org/10.1186/s40561-019-0097-y

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. (2014). Issues in researching self-regulated learning as patterns of events. In *Metacognition and Learning* (Vol. 9, Issue 2, pp. 229–237). Springer New York LLC. https://doi.org/10.1007/s11409-014-9113-3

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

Winne, P., & Hadwin, A. (1998). Studying as self-regulated learning. In *Metacognition in Educational Theory and Practice* (Vol. 93). Lawrence Erlbaum Associates Publishers.

Wise, A. F. (2014). Designing pedagogical interventions to support student use of learning analytics. *ACM International Conference Proceeding Series*, 203–211. https://doi.org/10.1145/2567574.2567588

Wise, A. F., & Shaffer, D. W. (2015). Why Theory Matters More than Ever in the Age of Big Data. *Journal of Learning Analytics*, *2*(2), 5–13. https://doi.org/10.18608/jla.2015.22.2

Wu, N., Zhang, L., Gao, Y., Zhang, M., Sun, X., & Feng, J. (2019). CLMS-Net: Dropout Prediction in MOOCs with Deep Learning. *Proceedings of the ACM Turing Celebration Conference - China*. https://doi.org/10.1145/3321408.3322848

Wu, S. Y., & Wang, S. M. (2020). Exploring the effects of gender grouping and the cognitive processing patterns of a Facebook-based online collaborative learning activity. *Interactive Learning Environments*. https://doi.org/10.1080/10494820.2020.1799026

Yang, K.-H., & Lu, B.-C. (2021). Towards the successful game-based learning: Detection and feedback to misconceptions is the key. *Computers & Education*, *160*, 104033. https://doi.org/https://doi.org/10.1016/j.compedu.2020.104033

Zhou, M., & Winne, P. H. (2012). Modeling academic achievement by self-reported versus traced goal orientation. *Learning and Instruction*, *22*(6), 413–419. https://doi.org/10.1016/j.learninstruc.2012.03.004

Zhou, M., Xu, Y., Nesbit, J., & Winne, P. (2010). Sequential Pattern Analysis of Learning Logs. In *Handbook of Educational Data Mining* (pp. 107–121). https://doi.org/10.1201/b10274-10

Zimmerman, B. J. (1990). Self-Regulated Learning and Academic Achievement: An Overview. *Educational Psychologist*, *25*(1), 3–17. https://doi.org/10.1207/s15326985ep2501_2

Appendix A. Supplementary Information for the Mapping Study Resources by Research Focuses

Research Focus / Insight / Technique	Works
At-risk student identification	
Collaboration	
Network Analysis	[31]
Other Prediction Models	[31]
Visualization Analysis	[31]
Learning indicators	
Frequent Sequence Analysis	[91]
Neural Network	[91]
Time and learning	
Network Analysis	[31]
Other Prediction Models	[31]
Statistical Analysis	[39]
Visualization Analysis	[31]
No learning focus outcome	
Cluster Analysis	[127], [151], [70], [7], [138]
Frequent Sequence Analysis	[138]
Neural Network	[168], [123]
Other Prediction Models	[71], [33], [93], [9], [151], [7], [13], [85], [88], [100], [101], [102], [123], [175]
Process Mining	[127], [70]
Statistical Analysis	[153], [103]
Visualization Analysis	[168], [127], [33], [70], [7]
Non-SRL learning indicators identification	
Collaboration	
Cluster Analysis	[75]
Frequent Sequence Analysis	[166]
Network Analysis	[86], [113]
Process Mining	[8], [164], [166]
Text Mining	[75]
Visualization Analysis	[75], [86]
Course design	
Cluster Analysis	[52]
Frequent Sequence Analysis	[52]
Process Mining	[158], [23], [169]
Qualitative Analysis	[98]
Statistical Analysis	[108], [23], [109]
Visualization Analysis	[108], [98], [109]
Feedback	

Research Focus / Insight / Technique	Works
Cluster Analysis	[10]
Process Mining	[10], [165], [169], [170]
Qualitative Analysis	[98]
Statistical Analysis	[99], [87], [165], [174]
Visualization Analysis	[10], [98]
Learning indicators	
Cluster Analysis	[47], [75], [84], [133]
Frequent Sequence Analysis	[47], [84], [133]
Network Analysis	[86]
Process Mining	[61], [8], [23], [24], [58], [59], [72], [84], [133], [135], [145], [144], [146], [148], [149], [150], [167], [170]
Statistical Analysis	[99], [157], [23], [51], [53], [66], [72], [83], [145], [161]
Text Mining	[75], [83], [148]
Visualization Analysis	[59], [75], [83], [86], [133], [135],
Time and learning	
Cluster Analysis	[154]
Statistical Analysis	[108], [126]
Visualization Analysis	[108], [126]
No learning focus outcome	
Other Prediction Models	[46]
Process Mining	[62]
Visualization Analysis	[62]
Exploring socio-dynamic	
Collaboration	
Cluster Analysis	[12], [74], [75]
Network Analysis	[76], [19], [74]
Process Mining	[18], [82], [171]
Qualitative Analysis	[78]
Statistical Analysis	[78], [82]
Text Mining	[76], [19], [74], [75]
Visualization Analysis	[76], [12], [75], [171]
Course design	
Cluster Analysis	[52]
Frequent Sequence Analysis	[52]
Feedback	
Cluster Analysis	[74]
Network Analysis	[74]
Statistical Analysis	[156]
Text Mining	[74]

Research Focus / Insight / Technique	Works
Learning indicators	
Cluster Analysis	[75]
Network Analysis	[112]
Process Mining	[56], [82], [171]
Qualitative Analysis	[78]
Statistical Analysis	[78], [82]
Text Mining	[75], [112]
Visualization Analysis	[56], [75], [112], [171]
Time and learning	
Cluster Analysis	[12]
Network Analysis	[19]
Statistical Analysis	[156]
Text Mining	[19]
Visualization Analysis	[12]
No learning focus outcome	
Statistical Analysis	[37]
Exploring SRL processes	
Collaboration	
Frequent Sequence Analysis	[173]
Process Mining	[43], [92], [142], [173]
Qualitative Analysis	[78]
Statistical Analysis	[32], [78], [142]
Visualization Analysis	[142]
Course design	
Cluster Analysis	[67], [122]
Frequent Sequence Analysis	[67]
Visualization Analysis	[67]
Feedback	
Cluster Analysis	[97]
Frequent Sequence Analysis	[97]
Process Mining	[97]
Visualization Analysis	[97]
Learning indicators	
Cluster Analysis	[97], [67], [96], [45], [28], [30], [57], [65], [132], [172]
Frequent Sequence Analysis	[97], [67], [96], [45], [28], [65], [147], [162], [172]
Network Analysis	[79]
Process Mining	[97], [43], [45], [17], [28], [36], [57], [60], [92], [115], [132], [162]
Qualitative Analysis	[78]
Statistical Analysis	[32], [36], [60], [78], [79], [141],

Research Focus / Insight / Technique	Works
Visualization Analysis	[97], [67], [17], [30], [57], [60], [65], [115], [141], [162]
Time and learning	
Cluster Analysis	[122]
Frequent Sequence Analysis	[173]
Process Mining	[173]
No learning focus outcome	
Other Prediction Models	[101]
Group emergence/ group comparison by performance	
Collaboration	
Frequent Sequence Analysis	[173]
Process Mining	[8], [22], [164], [173]
Visualization Analysis	[22]
Course design	
Statistical Analysis	[108]
Visualization Analysis	[108]
Learning indicators	
Cluster Analysis	[125], [64], [47], [172]
Frequent Sequence Analysis	[47], [162], [172]
Process Mining	[64], [8], [22], [24], [54], [72], [115], [144], [146], [148], [162]
Statistical Analysis	[25], [72], [141]
Text Mining	[148]
Visualization Analysis	[125], [64], [22], [115], [141], [162],
Time and learning	
Cluster Analysis	[2], [44]
Frequent Sequence Analysis	[173]
Process Mining	[2], [173]
Statistical Analysis	[108], [4]
Visualization Analysis	[108], [2]
No learning focus outcome	
Cluster Analysis	[127], [70], [3], [7]
Frequent Sequence Analysis	[3]
Other Prediction Models	[3], [7]
Process Mining	[127], [70]
Statistical Analysis	[38]
Visualization Analysis	[127], [70], [7]
Method or algorithm development	
Collaboration	
Network Analysis	[31], [113], [137]

Research Focus / Insight / Technique	Works
Other Prediction Models	[31], [49]
Process Mining	[128]
Text Mining	[42]
Visualization Analysis	[31]
Course design	
Cluster Analysis	[120]
Neural Network	[129], [140]
Other Prediction Models	[68]
Process Mining	[106], [73], [119], [176]
Qualitative Analysis	[63]
Statistical Analysis	[73], [81], [124], [176]
Visualization Analysis	[106], [129], [120], [176]
Feedback	
Cluster Analysis	[10], [97], [120]
Frequent Sequence Analysis	[97]
Other Prediction Models	[68]
Process Mining	[10], [97]
Statistical Analysis	[174]
Visualization Analysis	[10], [97], [120]
Learning indicators	
Cluster Analysis	[97], [64], [96], [27], [65], [84], [95], [132], [133]
Frequent Sequence Analysis	[97], [91], [96], [65], [84], [95], [133], [160]
Network Analysis	[131], [137]
Neural Network	[80], [91]
Other Prediction Models	[29], [107], [15]
Process Mining	[97], [64], [131], [17], [27], [73], [84], [95], [114], [130], [132], [133], [160], [176]
Statistical Analysis	[5], [157], [51], [73], [176]
Text Mining	[80], [42]
Visualization Analysis	[5], [97], [64], [131], [107], [48], [15], [17], [27], [41], [65], [95], [114], [133], [176]
Time and learning	
Cluster Analysis	[2], [152], [136]
Frequent Sequence Analysis	[20]
Network Analysis	[31], [152], [134]
Other Prediction Models	[31], [29], [117]
Process Mining	[2], [152], [155], [6], [20]
Qualitative Analysis	[63]
Statistical Analysis	[5], [124]

Research Focus / Insight / Technique	Works
Visualization Analysis	[31], [5], [2], [152], [6], [41],
No learning focus outcome	
Cluster Analysis	[3], [7], [11], [50], [90], [118], [138],
Frequent Sequence Analysis	[16], [69], [3], [138]
Neural Network	[168], [34], [55], [163], [1], [105],
Other Prediction Models	[139], [71], [77], [33], [93], [104], [9], [26], [3], [7], [40], [46], [111], [100], [101], [143]
Process Mining	[14], [94], [11], [50], [89], [118],
Statistical Analysis	[110], [116], [159], [121], [35]
Text Mining	[69]
Visualization Analysis	[14], [139], [168], [33], [34], [116], [105], [26], [7], [11], [50], [118]
Time to intervention	
Collaboration	
Network Analysis	[31]
Other Prediction Models	[31]
Visualization Analysis	[31]
Learning indicators	
Other Prediction Models	[29]
Time and learning	
Network Analysis	[31]
Other Prediction Models	[31], [29]
Statistical Analysis	[21]
Visualization Analysis	[31]

Appendix B. Supplementary information for the Mapping Study Resources by Learning Insights

Insight / Research Focus / Technique	Works
Collaboration	
At-risk student identification	
Network Analysis	[31]
Other Prediction Models	[31]
Visualization Analysis	[31]
Exploring socio-dynamic	
Cluster Analysis	[12], [74], [75]
Exploring SRL processes	
Frequent Sequence Analysis	[173]
Process Mining	[43], [92], [142], [173]
Qualitative Analysis	[78]
Statistical Analysis	[32], [78], [142]
Visualization Analysis	[142]
Group emergence/ group comparison by performance	
Frequent Sequence Analysis	[173]
Process Mining	[8], [22], [164], [173]
Visualization Analysis	[22]
Method or algorithm development	
Network Analysis	[31], [113], [137]
Other Prediction Models	[31], [49]
Process Mining	[128]
Text Mining	[42]
Visualization Analysis	[31]
Non-SRL learning indicators identification	
Cluster Analysis	[75]
Frequent Sequence Analysis	[166]
Network Analysis	[86], [113]
Process Mining	[8], [164], [166]
Text Mining	[75]
Visualization Analysis	[75], [86]
Time to intervention	
Network Analysis	[31]
Other Prediction Models	[31]
Visualization Analysis	[31]
Course design	
Exploring socio-dynamic	
Cluster Analysis	[52]
Frequent Sequence Analysis	[52]
Exploring SRL processes	
Cluster Analysis	[67], [122]
Frequent Sequence Analysis	[67]
Visualization Analysis	[67]
Group emergence/ group comparison by performance	

Insight / Research Focus / Technique	Works
Statistical Analysis	[108]
Visualization Analysis	[108]
Method or algorithm development	
Cluster Analysis	[120]
Neural Network	[129], [140]
Other Prediction Models	[68]
Process Mining	[106], [73], [119], [176]
Qualitative Analysis	[63]
Statistical Analysis	[73], [81], [124], [176]
Visualization Analysis	[106], [129], [120], [176]
Non-SRL learning indicators identification	
Cluster Analysis	[52]
Frequent Sequence Analysis	[52]
Process Mining	[158], [23], [169]
Qualitative Analysis	[98]
Statistical Analysis	[108], [23], [109]
Visualization Analysis	[108], [98], [109]
Feedback	
Exploring socio-dynamic	
Cluster Analysis	[74]
Network Analysis	[74]
Statistical Analysis	[156]
Text Mining	[74]
Exploring SRL processes	
Cluster Analysis	[97]
Frequent Sequence Analysis	[97]
Process Mining	[97]
Visualization Analysis	[97]
Method or algorithm development	
Cluster Analysis	[10], [97], [120]
Frequent Sequence Analysis	[97]
Other Prediction Models	[68]
Process Mining	[10], [97]
Statistical Analysis	[174]
Visualization Analysis	[10], [97], [120]
Non-SRL learning indicators identification	
Cluster Analysis	[10]

Insight / Research Focus / Technique	Works
Process Mining	[10], [165], [169], [170]
Qualitative Analysis	[98]
Statistical Analysis	[99], [87], [165], [174]
Visualization Analysis	[10], [98]
Learning indicators	
At-risk student identification	
Frequent Sequence Analysis	[91]
Neural Network	[91]
Exploring socio-dynamic	
Cluster Analysis	[75]
Network Analysis	[112]
Process Mining	[56], [82], [171]
Qualitative Analysis	[78]
Statistical Analysis	[78], [82]
Text Mining	[75], [112]
Visualization Analysis	[56], [75], [112], [171]
Exploring SRL processes	
Cluster Analysis	[97], [67], [96], [45], [28], [30], [57], [65], [132], [172]
Frequent Sequence Analysis	[97], [67], [96], [45], [28], [65], [147], [162], [172]
Network Analysis	[79]
Process Mining	[97], [43], [45], [17], [28], [36], [57], [60], [92], [115], [132], [162]
Qualitative Analysis	[78]
Statistical Analysis	[32], [36], [60], [78], [79], [141]
Visualization Analysis	[97], [67], [17], [30], [57], [60], [65], [115], [141], [162]
Group emergence/ group comparison by performance	
Cluster Analysis	[125], [64], [47], [172]
Frequent Sequence Analysis	[47], [162], [172]
Process Mining	[64], [8], [22], [24], [54], [72], [115], [144], [146], [148], [162]
Statistical Analysis	[25], [72], [141]
Text Mining	[148]
Visualization Analysis	[125], [64], [22], [115], [141], [162]
Method or algorithm development	
Cluster Analysis	[97], [64], [96], [27], [65], [84], [95], [132], [133]
Frequent Sequence Analysis	[97], [91], [96], [65], [84], [95], [133], [160]
Network Analysis	[131], [137]
Neural Network	[80], [91]
Other Prediction Models	[29], [107], [15]

Insight / Research Focus / Technique	Works
Process Mining	[97], [64], [131], [17], [27], [73], [84], [95], [114], [130], [132], [133], [160], [176]
Statistical Analysis	[5], [157], [51], [73], [176]
Text Mining	[80], [42]
Visualization Analysis	[5], [97], [64], [131], [107], [48], [15], [17], [27], [41], [65], [95], [114], [133], [176]
Non-SRL learning indicators identification	
Cluster Analysis	[47], [75], [84], [133]
Frequent Sequence Analysis	[47], [84], [133]
Network Analysis	[86]
Process Mining	[61], [8], [23], [24], [58], [59], [72], [84], [133], [135], [145], [144], [146], [148], [149], [150], [167], [170]
Statistical Analysis	[99], [157], [23], [51], [53], [66], [72], [83], [145], [161]
Text Mining	[75], [83], [148]
Visualization Analysis	[59], [75], [83], [86], [133], [135]
Time to intervention	
Other Prediction Models	[29]
No learning focus outcome	
At-risk student identification	
Cluster Analysis	[127], [151], [70], [7], [138]
Frequent Sequence Analysis	[138]
Neural Network	[168], [123]
Other Prediction Models	[71], [33], [93], [9], [151], [7], [13], [85], [88], [100], [101], [102], [123], [175]
Process Mining	[127], [70]
Statistical Analysis	[153], [103]
Visualization Analysis	[168], [127], [33], [70], [7]
Exploring socio-dynamic	
Statistical Analysis	[37]
Exploring SRL processes	
Other Prediction Models	[101]
Group emergence/ group comparison by performance	
Cluster Analysis	[127], [70], [3], [7]
Frequent Sequence Analysis	[3]
Other Prediction Models	[3], [7]
Process Mining	[127], [70]
Statistical Analysis	[38]
Visualization Analysis	[127], [70], [7]

Insight / Research Focus / Technique	Works
Method or algorithm development	
Cluster Analysis	[3], [7], [11], [50], [90], [118], [138]
Frequent Sequence Analysis	[16], [69], [3], [138]
Neural Network	[168], [34], [55], [163], [1], [105]
Other Prediction Models	[139], [71], [77], [33], [93], [104], [9], [26], [3], [7], [40], [46], [111], [100], [101], [143]
Process Mining	[14], [94], [11], [50], [89], [118]
Statistical Analysis	[110], [116], [159], [121], [35]
Text Mining	[69]
Visualization Analysis	[14], [139], [168], [33], [34], [116], [105], [26], [7], [11], [50], [118]
Non-SRL learning indicators identification	
Other Prediction Models	[46]
Process Mining	[62]
Visualization Analysis	[62]
Time and learning	
At-risk student identification	
Network Analysis	[31]
Other Prediction Models	[31]
Statistical Analysis	[39]
Visualization Analysis	[31]
Exploring socio-dynamic	
Cluster Analysis	[12]
Network Analysis	[19]
Statistical Analysis	[156]
Text Mining	[19]
Visualization Analysis	[12]
Exploring SRL processes	
Cluster Analysis	[122]
Frequent Sequence Analysis	[173]
Process Mining	[173]
Group emergence/ group comparison by performance	
Cluster Analysis	[2], [44]
Frequent Sequence Analysis	[173]
Process Mining	[2], [173]
Statistical Analysis	[108], [4]
Visualization Analysis	[108], [2]
Method or algorithm development	

Insight / Research Focus / Technique	Works
Cluster Analysis	[2], [152], [136]
Frequent Sequence Analysis	[20]
Network Analysis	[31], [152], [134]
Other Prediction Models	[31], [29], [117]
Process Mining	[2], [152], [155], [6], [20]
Qualitative Analysis	[63]
Statistical Analysis	[5], [124]
Visualization Analysis	[31], [5], [2], [152], [6], [41]
Non-SRL learning indicators identification	
Cluster Analysis	[154]
Statistical Analysis	[108], [126]
Visualization Analysis	[108], [126]
Time to intervention	
Network Analysis	[31]
Other Prediction Models	[31], [29]
Statistical Analysis	[21]
Visualization Analysis	[31]

WORKS ANALYZED

[1] Ghodai Abdelrahman and Qing Wang. 2019. Knowledge Tracing with Sequential Key-Value Memory Networks. In *Proceedings of the 42nd International ACM SIGIR Conference on Research and Development in Information Retrieval (SIGIR'19)*. Association for Computing Machinery, New York, NY, USA, 175–184. https://doi.org/10.1145/3331184.3331195

[2] Gökhan Akçapinar, Mei-Rong Alice Chen, Rwitajit Majumdar, Brendan Flanagan, and Hiroaki Ogata. 2020. Exploring Student Approaches to Learning through Sequence Analysis of Reading Logs. In *Proceedings of the Tenth International Conference on Learning Analytics Knowledge (LAK '20)*. Association for Computing Machinery, New York, NY, USA, 106–111. https://doi.org/10.1145/3375462.3375492

[3] Kamil Akhuseyinoglu and Peter Brusilovsky. 2021. Data-driven modeling of learners' individual differences for predicting engagement and success in online learning. *UMAP 2021 - Proceedings of the 29th ACM Conference on User Modeling, Adaptation and Personalization* 12, 21 (jun 2021), 201–212. https://doi.org/10.1145/3450613.3456834

[4] Abeer AlJarrah, Michael K Thomas, and Mohamed Shehab. 2018. Investigating temporal access in a flipped classroom: procrastination persists. *International Journal of Educational Technology in Higher Education* 15, 1 (2018), 1–18.

[5] Laura K Allen, Cecile Perret, Aaron Likens, and Danielle S McNamara. 2017. What'd You Say Again? Recurrence Quantification Analysis as a Method for Analyzing the Dynamics of Discourse in a Reading Strategy Tutor. In *Proceedings of the Seventh International Learning Analytics Knowledge Conference (LAK '17)*. Association for Computing Machinery, New York, NY, USA, 373–382. https://doi.org/10.1145/3027385.3027445

[6] Alejandro Andrade, Joshua A. Danish, and Adam V. Maltese. 2017. A Measurement Model of Gestures in an Embodied Learning Environment: Accounting for Temporal Dependencies. *Journal of Learning Analytics* 4, 3 (dec 2017), 18–45. https://doi.org/10.18608/jla.2017.43.3

[7] Raheela Asif, Agathe Merceron, Syed Abbas Ali, and Najmi Ghani Haider. 2017. Analyzing undergraduate students' performance using educational data mining. *Computers and Education* 113 (oct 2017), 177–194. https://doi.org/10.1016/j.compedu.2017.05.007

[8] Arif Bakla. 2018. Learner-generated materials in a flipped pronunciation class: A sequential explanatory mixed-methods study. *Computers Education* 125 (2018), 14–38. https://doi.org/10.1016/j.compedu.2018.05.017

[9] Jonathan Bassen, Bharathan Balaji, Michael Schaarschmidt, Candace Thille, Jay Painter, Dawn Zimmaro, Alex Games, Ethan Fast, and John C Mitchell. 2020. Reinforcement Learning for the Adaptive Scheduling of Educational Activities. In *Proceedings of the 2020 CHI Conference on Human Factors in Computing Systems (CHI '20)*. Association for Computing Machinery, New York, NY, USA, 1–12. https://doi.org/10.1145/3313831.3376518

[10] Mina Shirvani Boroujeni and Pierre Dillenbourg. 2018. Discovery and Temporal Analysis of Latent Study Patterns in MOOC Interaction Sequences. In *Proceedings of the 8th International Conference on Learning Analytics and Knowledge (LAK '18)*. Association for Computing Machinery, New York, NY, USA, 206–215. https://doi.org/10.1145/3170358.3170388

[11] Mina Shirvani Boroujeni and Pierre Dillenbourg. 2019. Discovery and temporal analysis of MOOC study patterns. *Journal of Learning Analytics* 6, 1 (apr 2019), 16–33. https://doi.org/10.18608/jla.2019.61.2

[12] Mina Shirvani Boroujeni, Tobias Hecking, H Ulrich Hoppe, and Pierre Dillenbourg. 2017. Dynamics of MOOC Discussion Forums. In *Proceedings of the Seventh International Learning Analytics Knowledge Conference (LAK '17)*. Association for Computing Machinery, New York, NY, USA, 128–137. https://doi.org/10.1145/3027385.3027391

[13] Michael Brown, R. Matthew DeMonbrun, and Stephanie Teasley. 2018. Taken Together: Conceptualizing Students' Concurrent Course Enrollment across the Post-Secondary Curriculum using temporal analytics. *Journal of Learning Analytics* 5, 3 (dec 2018), 60–72. https://doi.org/10.18608/jla.2018.53.5

[14] Marco Cameranesi, Claudia Diamantini, Laura Genga, and Domenico Potena. 2017. Students' Careers Analysis: A Process Mining Approach. In *Proceedings of the 7th International Conference on Web Intelligence, Mining and Semantics (WIMS '17)*. Association for Computing Machinery, New York, NY, USA. https://doi.org/10.1145/3102254.3102270

[15] Olga Caprotti. 2017. Shapes of Educational Data in an Online Calculus Course. *Journal of Learning Analytics* 4, 2 (jul 2017), 76–90. https://doi.org/10.18608/jla.2017.42.8

[16] Adam Scott Carter and Christopher David Hundhausen. 2017. Using Programming Process Data to Detect Differences in Students' Patterns of Programming. In *Proceedings of the 2017 ACM SIGCSE Technical Symposium on Computer Science Education (SIGCSE '17)*. Association for Computing Machinery, New York, NY, USA, 105–110. https://doi.org/10.1145/3017680.3017785

[17] Rebeca Cerezo, Alejandro Bogarín, María Esteban, and Cristóbal Romero. 2020. Process mining for self-regulated learning assessment in e-learning. *Journal of computing in higher education* 32, 1 (2020), 74–88.

[18] Chia-Jung Chang, Ming-Hua Chang, Bing-Cheng Chiu, Chen-Chung Liu, Shih-Hsun Fan Chiang, Cai-Ting Wen, Fu-Kwun Hwang, Ying-Tien Wu, Po-Yao Chao, Chia-Hsi Lai, Su-Wen Wu, Chih-Kang Chang, and Wenli Chen. 2017. An analysis of student collaborative problem solving activities mediated by collaborative simulations. *Computers Education* 114 (2017), 222–235. https://doi.org/10.1016/j.compedu.2017.07.008

[19] Bodong Chen and Oleksandra Poquet. 2020. Socio-Temporal Dynamics in Peer Interaction Events. In *Proceedings of the Tenth International Conference on Learning Analytics Knowledge (LAK '20)*. Association for Computing Machinery, New York, NY, USA, 203–208. https://doi.org/10.1145/3375462.3375535

[20] Bodong Chen, Monica Resendes, Ching Sing Chai, and Huang Yao Hong. 2017. Two tales of time: uncovering the significance of sequential patterns among contribution types in knowledge-building discourse. *Interactive Learning Environments* 25, 2 (feb 2017), 162–175. https://doi.org/10.1080/10494820.2016.1276081

[21] Lujie Karen Chen. 2021. Timing of support in one-on-one math problem solving coaching: A survival analysis approach with multimodal data. *ACM International Conference Proceeding Series* (2021), 553–558. https://doi.org/10.1145/3448139.3448197

[22] Hercy N.H. Cheng, Zhi Liu, Jianwen Sun, Sanya Liu, and Zongkai Yang. 2017. Unfolding online learning behavioral patterns and their temporal changes of college students in SPOCs. *Interactive Learning Environments* 25, 2 (feb 2017), 176–188. https://doi.org/10.1080/10494820.2016.1276082

[23] Kun-Hung Cheng and Chin-Chung Tsai. 2019. A case study of immersive virtual field trips in an elementary classroom: Students' learning experience and teacher-student interaction behaviors. *Computers Education* 140 (2019), 103600. https://doi.org/10.1016/j.compedu.2019.103600

[24] Guo Li Chiou, Chung Yuan Hsu, and Meng Jung Tsai. 2019. Exploring how students interact with guidance in a physics simulation: evidence from eye-movement and log data analyses. *Interactive Learning Environments* (2019). https://doi.org/10.1080/10494820.2019.1664596

[25] MIng Ming Chiu. 2018. Statistically Modelling Effects of Dynamic Processes on Outcomes: An Example of Discourse Sequences and Group Solutions. *Journal of Learning Analytics* 5, 1 (apr 2018), 75. https://doi.org/10.18608/jla.2018.51.6

[26] Youngduck Choi, Youngnam Lee, Junghyun Cho, Jineon Baek, Byungsoo Kim, Yeongmin Cha, Dongmin Shin, Chan Bae, and Jaewe Heo. 2020. Towards an Appropriate Query, Key, and Value Computation for Knowledge Tracing. In *Proceedings of the Seventh ACM Conference on Learning @ Scale (L@S '20)*. Association for Computing Machinery, New York, NY, USA, 341–344. https://doi.org/10.1145/3386527.3405945

[27] David Codish, Eyal Rabin, and Gilad Ravid. 2019. User behavior pattern detection in unstructured processes – a learning management system case study. *Interactive Learning Environments* 27, 5-6 (aug 2019), 699–725. https://doi.org/10.1080/10494820.2019.1610456

[28] Matt Crosslin, Kimberly Breuer, Nikola Milikić, and Justin T. Dellinger. 2021. Understanding student learning pathways in traditional online history courses: utilizing process mining analysis on clickstream data. *Journal of Research in Innovative Teaching Learning* 14, 3 (nov 2021), 399–414. https://doi.org/10.1108/JRIT-03-2021-0024

[29] Steven Dang, Michael Yudelson, and Kenneth R Koedinger. 2017. Detecting Diligence with Online Behaviors on Intelligent Tutoring Systems. In *Proceedings of the Fourth (2017) ACM Conference on Learning @ Scale (L@S '17)*. Association for Computing Machinery, New York, NY, USA, 51–59. https://doi.org/10.1145/3051457.3051470

[30] Paula G de Barba, Donia Malekian, Eduardo A Oliveira, James Bailey, Tracii Ryan, and Gregor Kennedy. 2020. The importance and meaning of session behaviour in a MOOC. *Computers Education* 146 (2020), 103772. https://doi.org/10.1016/j.compedu.2019.103772

[31] Nicholas Diana, Michael Eagle, John Stamper, Shuchi Grover, Marie Bienkowski, and Satabdi Basu. 2017. An Instructor Dashboard for Real-Time Analytics in Interactive Programming Assignments. In *Proceedings of the Seventh International Learning Analytics Knowledge Conference (LAK '17)*. Association for Computing Machinery, New York, NY, USA, 272–279. https://doi.org/10.1145/3027385.3027441

[32] Muhterem Dindar, Jonna Malmberg, Sanna Järvelä, Eetu Haataja, and Paul A. Kirschner. 2020. Matching self-reports with electrodermal activity data: Investigating temporal changes in self-regulated learning. *Education and Information Technologies* 25, 3 (may 2020), 1785–1802. https://doi.org/10.1007/s10639-019-10059-5

[33] Mucong Ding, Yanbang Wang, Erik Hemberg, and Una-May O'Reilly. 2019. Transfer Learning Using Representation Learning in Massive Open Online Courses. In *Proceedings of the 9th International Conference on Learning Analytics Knowledge (LAK19)*. Association for Computing Machinery, New York, NY, USA, 145–154. https://doi.org/10.1145/3303772.3303794

[34] Mucong Ding, Kai Yang, Dit-Yan Yeung, and Ting-Chuen Pong. 2019. Effective Feature Learning with Unsupervised Learning for Improving the Predictive Models in Massive Open Online Courses. In *Proceedings of the 9th International Conference on Learning Analytics Knowledge (LAK19)*. Association for Computing Machinery, New York, NY, USA, 135–144. https://doi.org/10.1145/3303772.3303795

[35] Tenzin Doleck, Susanne P. Lajoie, and Paul Bazelais. 2019. Social networking and academic performance: A longitudinal perspective. *Education and Information Technologies* 24, 2 (mar 2019), 1545–1561. https://doi.org/10.1007/s10639-018-9843-y

[36] Cesar Dominguez, Francisco J. Garcia-Izquierdo, Arturo Jaime, Beatriz Perez, Angel Luis Rubio, and Maria A. Zapata. 2021. Using Process Mining to Analyze Time Distribution of Self-Assessment and Formative Assessment Exercises on an Online Learning Tool. *IEEE Transactions on Learning Technologies* 14, 5 (2021), 709–722. https://doi.org/10.1109/TLT.2021.3119224

[37] Nia M M Dowell, Christopher Brooks, Vitomir Kovanović, Srećko Joksimović, and Dragan Gašević. 2017. The Changing Patterns of MOOC Discourse. In *Proceedings of the Fourth (2017) ACM Conference on Learning @ Scale (L@S '17)*. Association for Computing Machinery, New York, NY, USA, 283–286. https://doi.org/10.1145/3051457.3054005

[38] Xu Du, Mingyan Zhang, Brett E. Shelton, and Jui Long Hung. 2019. Learning anytime, anywhere: a spatio-temporal analysis for online learning. *Interactive Learning Environments* (2019). https://doi.org/10.1080/10494820.2019.1633546

[39] Rebecca L Edwards, Sarah K Davis, Allyson F Hadwin, and Todd M Milford. 2017. Using Predictive Analytics in a Self-Regulated Learning University Course to Promote Student Success. In *Proceedings of the Seventh International Learning Analytics Knowledge Conference (LAK '17)*. Association for Computing Machinery, New York, NY, USA, 556–557. https://doi.org/10.1145/3027385.3029455

[40] Houssam El Aouifi, Mohamed El Hajji, Youssef Es-Saady, and Hassan Douzi. 2021. Predicting learner's performance through video sequences viewing behavior analysis using educational data-mining. *Education and Information Technologies* 26, 5 (sep 2021), 5799–5814. https://doi.org/10.1007/S10639-021-10512-4/TABLES/7

[41] Volkmar P. Engerer. 2020. Implementing dynamicity in research designs for collaborative digital writing. *Education and Information Technologies* (may 2020). https://doi.org/10.1007/s10639-020-10365-3

[42] Volkmar P. Engerer. 2021. Temporality revisited: Dynamicity issues in collaborative digital writing research. *Education and Information Technologies* 26, 1 (jan 2021), 339–370. https://doi.org/10.1007/s10639-020-10262-9

[43] Erkan Er, Cristina Villa-Torrano, Yannis Dimitriadis, Dragan Gasevic, Miguel L Bote-Lorenzo, Juan I Asensio-Pérez, Eduardo Gómez-Sánchez, and Alejandra Mart´Monés. 2021. Theory-Based Learning Analytics to Explore Student Engagement Patterns in a Peer Review Activity. In *LAK21: 11th International Learning Analytics and Knowledge Conference (LAK21)*. Association for Computing Machinery, New York, NY, USA, 196–206. https://doi.org/10.1145/3448139.3448158

[44] Leon Fadljević, Katharina Maitz, Dominik Kowald, Viktoria Pammer-Schindler, and Barbara Gasteiger-Klicpera. 2020. Slow is Good: The Effect of Diligence on Student Performance in the Case of an Adaptive Learning System for Health Literacy. In *Proceedings of the Tenth International Conference on Learning Analytics Knowledge (LAK '20)*. Association for Computing Machinery, New York, NY, USA, 112–117. https://doi.org/10.1145/3375462.3375502

[45] Yizhou Fan, John Saint, Shaveen Singh, Jelena Jovanovic, and Dragan Gašević. 2021. A Learning Analytic Approach to Unveiling Self-Regulatory Processes in Learning Tactics. In *LAK21: 11th International Learning Analytics and Knowledge Conference (LAK21)*. Association for Computing Machinery, New York, NY, USA, 184–195. https://doi.org/10.1145/3448139.3448211

[46] Somayeh Fatahi, Faezeh Shabanali-Fami, and Hadi Moradi. 2018. An empirical study of using sequential behavior pattern mining approach to predict learning styles. *Education and Information Technologies* 23, 4 (jul 2018), 1427–1445. https://doi.org/10.1007/s10639-017-9667-1

[47] Dragan Gasevic, Jelena Jovanovic, Abelardo Pardo, and Shane Dawson. 2017. Detecting Learning Strategies with Analytics: Links with Self-reported Measures and Academic Performance. *Journal of Learning Analytics* 4, 2 (jul 2017), 113–128. https://doi.org/10.18608/jla.2017.42.10

[48] Manuel J Gomez, José A Ruipérez-Valiente, Pedro A Martinez, and Yoon Jeon Kim. 2020. Exploring the Affordances of Sequence Mining in Educational Games. In *Eighth International Conference on Technological Ecosystems for Enhancing Multiculturality (TEEM'20)*. Association for Computing Machinery, New York, NY, USA, 648–654. https://doi.org/10.1145/3434780.3436562

[49] Zhen Guo, Zhe Zhang, and Munindar Singh. 2020. In Opinion Holders' Shoes: Modeling Cumulative Influence for View Change in Online Argumentation. In *Proceedings of The Web Conference 2020 (WWW '20)*. Association for Computing Machinery, New York, NY, USA, 2388–2399. https://doi.org/10.1145/3366423.3380302

[50] Christian Hansen, Casper Hansen, Niklas Hjuler, Stephen Alstrup, and Christina Lioma. 2017. Sequence modelling for analysing student interaction with educational systems. In *Proceedings of the 10th International Conference on Educational Data Mining, EDM 2017*. International Educational Data Mining Society, 232–237. arXiv:1708.04164

[51] Qiwei He, Francesca Borgonovi, and Marco Paccagnella. 2021. Leveraging process data to assess adults' problem-solving skills: Using sequence mining to identify behavioral patterns across digital tasks. *Computers and Education* 166 (jun 2021), 104170. https://doi.org/10.1016/J.COMPEDU.2021.104170

[52] Tobias Hecking, Irene Angelica Chounta, and H. Ulrich Hoppe. 2017. Role Modelling in MOOC Discussion Forums. *Journal of Learning Analytics* 4, 1 (mar 2017), 85–116. https://doi.org/10.18608/jla.2017.41.6

[53] Vo Ngoc Hoi and Ho Le Hang. 2021. Understanding students' behavioural intention to use facebook as a supplementary learning platform: A mixed methods approach. *Education and Information Technologies* 26, 5 (sep 2021), 5991–6011. https://doi.org/10.1007/S10639-021-10565-5/TABLES/4

[54] Chung Yuan Hsu, Guo Li Chiou, and Meng Jung Tsai. 2019. Visual behavior and self-efficacy of game playing: an eye movement analysis. *Interactive Learning Environments* 27, 7 (2019), 942–952. https://doi.org/10.1080/10494820.2018.1504309

[55] Qian Hu and Huzefa Rangwala. 2019. Reliable Deep Grade Prediction with Uncertainty Estimation. In *Proceedings of the 9th International Conference on Learning Analytics Knowledge (LAK19)*. Association for Computing Machinery, New York, NY, USA, 76–85. https://doi.org/10.1145/3303772.3303802

[56] Chang-Qin Huang, Zhong-Mei Han, Ming-Xi Li, Morris Siu-yung Jong, and Chin-Chung Tsai. 2019. Investigating students' interaction patterns and dynamic learning sentiments in online discussions. *Computers Education* 140 (2019), 103589. https://doi.org/10.1016/j.compedu.2019.05.015

[57] Lingyun Huang and Susanne P Lajoie. 2021. Process analysis of teachers' self-regulated learning patterns in technological pedagogical content knowledge development. *Computers Education* 166 (2021), 104169. https://doi.org/10.1016/j.compedu.2021.104169

[58] Gwo-Jen Hwang and Chih-Hung Chen. 2017. Influences of an inquiry-based ubiquitous gaming design on students' learning achievements, motivation, behavioral patterns, and tendency towards critical thinking and problem solving. *British Journal of Educational Technology* 48, 4 (2017), 950–971. https://doi.org/10.1111/bjet.12464

[59] Gwo-Jen Hwang, Ting-Chia Hsu, Chiu-Lin Lai, and Ching-Jung Hsueh. 2017. Interaction of problem-based gaming and learning anxiety in language students' English listening performance and progressive behavioral patterns. *Computers Education* 106 (2017), 26–42. https://doi.org/10.1016/j.compedu.2016.11.010

[60] Gwo-Jen Hwang, Sheng-Yuan Wang, and Chiu-Lin Lai. 2021. Effects of a social regulation-based online learning framework on students' learning achievements and behaviors in mathematics. *Computers Education* 160 (2021), 104031. https://doi.org/10.1016/j.compedu.2020.104031

[61] Allan Jeong, Haiying Li, and Andy Jiaren Pan. 2017. A sequential analysis of responses in online debates to postings of students exhibiting high versus low grammar and spelling errors. *Educational Technology Research and Development* 65, 5 (2017), 1175–1194. http://www.jstor.org/stable/45018721

[62] Laipeng Jin and Dongchuan Yu. 2019. Characteristics of Visual Attention for the Assessment of Conceptual Change: An Eye-Tracking Study. In *Proceedings of the 10th International Conference on E-Education, E-Business, E-Management and E-Learning (IC4E '19)*. Association for Computing Machinery, New York, NY, USA, 158–162. https://doi.org/10.1145/3306500.3306584

[63] Aditya Johri. 2018. How FLOSS Participation Supports Lifelong Learning and Working: Apprenticeship Across Time and Spatialities. In *Proceedings of the 14th International Symposium on Open Collaboration (OpenSym '18)*. Association for Computing Machinery, New York, NY, USA. https://doi.org/10.1145/3233391.3233541

[64] Jelena Jovanović, Shane Dawson, Srećko Joksimović, and George Siemens. 2020. Supporting Actionable Intelligence: Reframing the Analysis of Observed Study Strategies. In *Proceedings of the Tenth International Conference on Learning Analytics Knowledge (LAK '20)*. Association for Computing Machinery, New York, NY, USA, 161–170. https://doi.org/10.1145/3375462.3375474

[65] Jelena Jovanović, Dragan Gašević, Shane Dawson, Abelardo Pardo, and Negin Mirriahi. 2017. Learning analytics to unveil learning strategies in a flipped classroom. *Internet and Higher Education* 33 (apr 2017), 74–85. https://doi.org/10.1016/j.iheduc.2017.02.001

[66] Jelena Jovanović, Dragan Gašević, Abelardo Pardo, Shane Dawson, and Alexander Whitelock-Wainwright. 2019. Introducing meaning to clicks: Towards traced-measures of self-efficacy and cognitive load. In *Proceedings of the 9th International Conference on Learning Analytics Knowledge - LAK19 (LAK19)*. ACM Press, New York, New York, USA, 511–520. https://doi.org/10.1145/3303772.3303782

[67] Fatemeh Salehian Kia, Stephanie D Teasley, Marek Hatala, Stuart A Karabenick, and Matthew Kay. 2020. How Patterns of Students Dashboard Use Are Related to Their Achievement and Self-Regulatory Engagement. In *Proceedings of the Tenth International Conference on Learning Analytics Knowledge (LAK '20)*. Association for Computing Machinery, New York, NY, USA, 340–349. https://doi.org/10.1145/3375462.3375472

[68] Aleksandra Klašnja-Milićević, Boban Vesin, and Mirjana Ivanović. 2018. Social tagging strategy for enhancing e-learning experience. *Computers Education* 118 (2018), 166–181. https://doi.org/10.1016/j.compedu.2017.12.002

[69] Simon Knight, Roberto Martinez-Maldonado, Andrew Gibson, and Simon Buckingham Shum. 2017. Towards Mining Sequences and Dispersion of Rhetorical Moves in Student Written Texts. In *Proceedings of the Seventh International Learning Analytics Knowledge Conference (LAK '17)*. Association for Computing Machinery, New York, NY, USA, 228–232. https://doi.org/10.1145/3027385.3027433

[70] Mehmet Kokoç, Gökhan Akçapınar, Mohammad Nehal Hasnine, Mehmet Kokoc, Gokhan Akcapinar, and Mohammad Nehal Hasnine. 2021. Unfolding Students' Online Assignment Submission Behavioral Patterns using Temporal Learning Analytics. *Educational Technology Society* 24, 1 (2021), 223–235. https://www.jstor.org/stable/26977869

[71] Srijan Kumar, Xikun Zhang, and Jure Leskovec. 2019. Predicting Dynamic Embedding Trajectory in Temporal Interaction Networks. In *Proceedings of the 25th ACM SIGKDD International Conference on Knowledge Discovery Data Mining (KDD '19)*. Association for Computing Machinery, New York, NY, USA, 1269–1278. https://doi.org/10.1145/3292500.3330895

[72] Wei Chen Kuo and Ting Chia Hsu. 2020. Learning Computational Thinking Without a Computer: How Computational Participation Happens in a Computational Thinking Board Game. *Asia-Pacific Education Researcher* 29, 1 (feb 2020), 67–83. https://doi.org/10.1007/s40299-019-00479-9

[73] Joni Lämsä, Raija Hämäläinen, Pekka Koskinen, Jouni Viiri, and Joonas Mannonen. 2020. The potential of temporal analysis: Combining log data and lag sequential analysis to investigate temporal differences between scaffolded and non-scaffolded group inquiry-based learning processes. *Computers Education* 143 (2020), 103674. https://doi.org/10.1016/j.compedu.2019.103674

[74] Alwyn Vwen Yen Lee. 2021. Determining Quality and Distribution of Ideas in Online Classroom Talk using Learning Analytics and Machine Learning. *Educational Technology Society* 24, 1 (2021), 236–249. https://www.jstor.org/stable/26977870

[75] Alwyn Vwen Yen Lee and Seng Chee Tan. 2017. Promising Ideas for Collective Advancement of Communal Knowledge Using Temporal Analytics and Cluster Analysis. *Journal of Learning Analytics* 4, 3 (dec 2017), 76–101. https://doi.org/10.18608/jla.2017.43.5

[76] Alwyn Vwen Yen Lee and Seng Chee Tan. 2017. Temporal Analytics with Discourse Analysis: Tracing Ideas and Impact on Communal Discourse. In *Proceedings of the Seventh International Learning Analytics Knowledge Conference (LAK '17)*. Association for Computing Machinery, New York, NY, USA, 120–127. https://doi.org/10.1145/3027385.3027386

[77] Jinseok Lee and Dit-Yan Yeung. 2019. Knowledge Query Network for Knowledge Tracing: How Knowledge Interacts with Skills. In *Proceedings of the 9th International Conference on Learning Analytics Knowledge (LAK19)*. Association for Computing Machinery, New York, NY, USA, 491–500. https://doi.org/10.1145/3303772.3303786

[78] Lila Lee, Susanne P. Lajoie, Eric G. Poitras, Miriam Nkangu, and Tenzin Doleck. 2017. Co-regulation and knowledge construction in an online synchronous problem based learning setting. *Education and Information Technologies* 22, 4 (jul 2017), 1623–1650. https://doi.org/10.1007/s10639-016-9509-6

[79] Shan Li, Hanxiang Du, Wanli Xing, Juan Zheng, Guanhua Chen, and Charles Xie. 2020. Examining temporal dynamics of self-regulated learning behaviors in STEM learning: A network approach. *Computers Education* 158 (2020), 103987. https://doi.org/10.1016/j.compedu.2020.103987

[80] Aaron D Likens, Kathryn S McCarthy, Laura K Allen, and Danielle S McNamara. 2018. Recurrence Quantification Analysis as a Method for Studying Text Comprehension Dynamics. In *Proceedings of the 8th International Conference on Learning Analytics and Knowledge (LAK '18)*. Association for Computing Machinery, New York, NY, USA, 111–120. https://doi.org/10.1145/3170358.3170407

[81] Ran Liu, John C Stamper, and Jodi Davenport. 2018. A Novel Method for the In-Depth Multimodal Analysis of Student Learning Trajectories in Intelligent Tutoring Systems. *Journal of Learning Analytics* 5, 1 (apr 2018), 41–54. https://doi.org/10.18608/jla.2018.51.4

[82] Sannyuya Liu, Lingyun Kang, Zhi Liu, Jing Fang, Zongkai Yang, Jianwen Sun, Meiyi Wang, and Mengwei Hu. 2021. Computer-supported collaborative concept mapping: the impact of students' perceptions of collaboration on their knowledge understanding and behavioral patterns. *Interactive Learning Environments* (may 2021), 1–20. https://doi.org/10.1080/10494820.2021.1927115

[83] Zhi Liu, Chongyang Yang, Sylvio Rüdian, Sannyuya Liu, Liang Zhao, and Tai Wang. 2019. Temporal emotion-aspect modeling for discovering what students are concerned about in online course forums. *Interactive Learning Environments* 27, 5-6 (aug 2019), 598–627. https://doi.org/10.1080/10494820.2019.1610449

[84] Sonsoles López-pernas, Mohammed Saqr, and Olga Viberg. 2021. Putting It All Together: Combining Learning Analytics Methods and Data Sources to Understand Students' Approaches to Learning Programming. *Sustainability 2021, Vol. 13, Page 4825* 13, 9 (apr 2021), 4825. https://doi.org/10.3390/SU13094825

[85] Owen H T Lu, Anna Y Q Huang, Jeff C H Huang, Albert J Q Lin, Hiroaki Ogata, and Stephen J H Yang. 2018. Applying Learning Analytics for the Early Prediction of Students' Academic Performance in Blended Learning. *Educational technology society* 21, 2 (2018), 220–232.

[86] Kristine Lund, Mattieu Quignard, and David Williamson Shaffer. 2017. Gaining Insight by Transforming Between Temporal Representations of Human Interaction. *Journal of Learning Analytics* 4, 3 (dec 2017), 102–122. https://doi.org/10.18608/jla.2017.43.6

[87] Jiwen Luo and Tao Wang. 2020. Analyzing Students' Behavior in Blended Learning Environment for Programming Education. In *Proceedings of the 2020 The 2nd World Symposium on Software Engineering (WSSE 2020)*. Association for Computing Machinery, New York, NY, USA, 179–185. https://doi.org/10.1145/3425329.3425346

[88] Charles Lwande, Robert Oboko, and Lawrence Muchemi. 2021. Learner behavior prediction in a learning management system. *Education and Information Technologies* 26, 3 (may 2021), 2743–2766. https://doi.org/10.1007/S10639-020-10370-6/TABLES/9

[89] Martin Macak, Daniela Kruzelova, Stanislav Chren, and Barbora Buhnova. 2021. Using process mining for Git log analysis of projects in a software development course. *Education and Information Technologies* (may 2021), 1–31. https://doi.org/10.1007/s10639-021-10564-6

[90] Mohammad Javad Mahzoon, Mary Lou Maher, Omar Eltayeby, Wenwen Dou, and Kazjon Grace. 2018. A Sequence Data Model for Analyzing Temporal Patterns of Student Data. *Journal of Learning Analytics* 5, 1 (apr 2018), 55–74. https://doi.org/10.18608/jla.2018.51.5

[91] Donia Malekian, James Bailey, and Gregor Kennedy. 2020. Prediction of Students' Assessment Readiness in Online Learning Environments: The Sequence Matters. In *Proceedings of the Tenth International Conference on Learning Analytics Knowledge (LAK '20)*. Association for Computing Machinery, New York, NY, USA, 382–391. https://doi.org/10.1145/3375462.3375468

[92] Jonna Malmberg, Sanna Järvelä, and Hanna Järvenoja. 2017. Capturing temporal and sequential patterns of self-, co-, and socially shared regulation in the context of collaborative learning. *Contemporary Educational Psychology* 49 (apr 2017), 160–174. https://doi.org/10.1016/j.cedpsych.2017.01.009

[93] Rubén Manrique, Bernardo Pereira Nunes, Olga Marino, Marco Antonio Casanova, and Terhi Nurmikko-Fuller. 2019. An Analysis of Student Representation, Representative Features and Classification Algorithms to Predict Degree Dropout. In *Proceedings of the 9th International Conference on Learning Analytics Knowledge (LAK19)*. Association for Computing Machinery, New York, NY, USA, 401–410. https://doi.org/10.1145/3303772.3303800

[94] Jeffrey Matayoshi and Shamya Karumbaiah. 2021. Using Marginal Models to Adjust for Statistical Bias in the Analysis of State Transitions. In *LAK21: 11th International Learning Analytics and Knowledge Conference (LAK21)*. Association for Computing Machinery, New York, NY, USA, 449–455. https://doi.org/10.1145/3448139.3448182

[95] Wannisa Matcha, Dragan Gašević, Nora'ayu Ahmad Uzir, Jelena Jovanović, Abelardo Pardo, Lisa Lim, Jorge Maldonado-Mahauad, Sheridan Gentili, Mar Pérez-Sanagustín, and Yi-Shan Tsai. 2020. Analytics of Learning Strategies: Role of Course Design and Delivery Modality. *Journal of Learning*

Analytics 7, 2 (sep 2020), 45–71. https://doi.org/10.18608/jla.2020.72.3
[96] Wannisa Matcha, Dragan Gašević, Jelena Jovanović, Nora'ayu Ahmad Uzir, Chris W Oliver, Andrew Murray, and Danijela Gasevic. 2020. Analytics of Learning Strategies: The Association with the Personality Traits. In *Proceedings of the Tenth International Conference on Learning Analytics Knowledge (LAK '20)*. Association for Computing Machinery, New York, NY, USA, 151–160. https://doi.org/10.1145/3375462.3375534
[97] Wannisa Matcha, Dragan Gašević, Nora'Ayu Ahmad Uzir, Jelena Jovanović, and Abelardo Pardo. 2019. Analytics of Learning Strategies: Associations with Academic Performance and Feedback. In *Proceedings of the 9th International Conference on Learning Analytics Knowledge (LAK19)*. Association for Computing Machinery, New York, NY, USA, 461–470. https://doi.org/10.1145/3303772.3303787
[98] Douglas McHugh, Richard Feinn, Jeff McIlvenna, and Matt Trevithick. 2021. A Random Controlled Trial to Examine the Efficacy of Blank Slate: A Novel Spaced Retrieval Tool with Real-Time Learning Analytics. *Education sciences* 11, 3 (2021), 90.
[99] Ritayan Mitra and Pankaj Chavan. 2019. DEBE Feedback for Large Lecture Classroom Analytics. In *Proceedings of the 9th International Conference on Learning Analytics Knowledge (LAK19)*. Association for Computing Machinery, New York, NY, USA, 426–430. https://doi.org/10.1145/3303772.3303821
[100] Pedro Manuel Moreno-Marcos, Pedro J Muñoz-Merino, Carlos Alario-Hoyos, Iria Estévez-Ayres, and Carlos Delgado Kloos. 2018. Analysing the predictive power for anticipating assignment grades in a massive open online course. *Behaviour information technology* 37, 10-11 (2018), 1021–1036.
[101] Pedro Manuel Moreno-Marcos, Pedro J Muñoz-Merino, Jorge Maldonado-Mahauad, Mar Pérez-Sanagustín, Carlos Alario-Hoyos, and Carlos Delgado Kloos. 2020. Temporal analysis for dropout prediction using self-regulated learning strategies in self-paced MOOCs. *Computers Education* 145 (2020), 103728. https://doi.org/10.1016/j.compedu.2019.103728
[102] Ahmed Ali Mubarak, Han Cao, and Salah A.M. Ahmed. 2021. Predictive learning analytics using deep learning model in MOOCs' courses videos. *Education and Information Technologies* 26, 1 (jan 2021), 371–392. https://doi.org/10.1007/S10639-020-10273-6/TABLES/7
[103] Ahmed A. Mubarak, Han Cao, and Weizhen Zhang. 2020. Prediction of students' early dropout based on their interaction logs in online learning environment. *Interactive Learning Environments* (2020). https://doi.org/10.1080/10494820.2020.1727529
[104] Koki Nagatani, Qian Zhang, Masahiro Sato, Yan-Ying Chen, Francine Chen, and Tomoko Ohkuma. 2019. Augmenting Knowledge Tracing by Considering Forgetting Behavior. In *The World Wide Web Conference (WWW '19)*. Association for Computing Machinery, New York, NY, USA, 3101–3107. https://doi.org/10.1145/3308558.3313565
[105] Hiromi Nakagawa, Yusuke Iwasawa, and Yutaka Matsuo. 2019. Graph-Based Knowledge Tracing: Modeling Student Proficiency Using Graph Neural Network. In *IEEE/WIC/ACM International Conference on Web Intelligence (WI '19)*. Association for Computing Machinery, New York, NY, USA, 156–163. https://doi.org/10.1145/3350546.3352513
[106] Andres Neyem, Juan Diaz-Mosquera, Jorge Munoz-Gama, and Jaime Navon. 2017. Understanding Student Interactions in Capstone Courses to Improve Learning Experiences. In *Proceedings of the 2017 ACM SIGCSE Technical Symposium on Computer Science Education (SIGCSE '17)*. Association for Computing Machinery, New York, NY, USA, 423–428. https://doi.org/10.1145/3017680.3017716
[107] Quan Nguyen. 2020. Rethinking Time-on-Task Estimation with Outlier Detection Accounting for Individual, Time, and Task Differences. In *Proceedings of the Tenth International Conference on Learning Analytics Knowledge (LAK '20)*. Association for Computing Machinery, New York, NY, USA, 376–381. https://doi.org/10.1145/3375462.3375538
[108] Quan Nguyen, Michal Huptych, and Bart Rienties. 2018. Linking Students' Timing of Engagement to Learning Design and Academic Performance. In *Proceedings of the 8th International Conference on Learning Analytics and Knowledge (LAK '18)*. Association for Computing Machinery, New York, NY, USA, 141–150. https://doi.org/10.1145/3170358.3170398
[109] Quan Nguyen, Michal Huptych, and Bart Rienties. 2018. Using Temporal Analytics to Detect Inconsistencies Between Learning Design and Students' Behaviours. *Journal of Learning Analytics* 5, 3 (dec 2018), 120–135. https://doi.org/10.18608/jla.2018.53.8
[110] Klinsukon Nimkanjana and Suntorn Witosurapot. 2018. Video-Based Question Generation for Mobile Learning. In *Proceedings of the 2nd International Conference on Education and Multimedia Technology (ICEMT 2018)*. Association for Computing Machinery, New York, NY, USA, 5–8. https://doi.org/10.1145/3206129.3239427
[111] Jennifer K Olsen, Kshitij Sharma, Nikol Rummel, and Vincent Aleven. 2020. Temporal analysis of multimodal data to predict collaborative learning outcomes. *British Journal of Educational Technology* 51, 5 (2020), 1527–1547. https://doi.org/10.1111/bjet.12982
[112] Jun Oshima, Ritsuko Oshima, and Wataru Fujita. 2018. A Mixed-Methods Approach to Analyze Shared Epistemic Agency in Jigsaw Instruction at Multiple Scales of Temporality. *Journal of Learning Analytics* 5, 1 (apr 2018), 10–24. https://doi.org/10.18608/jla.2018.51.2
[113] Fan Ouyang. 2021. Using Three Social Network Analysis Approaches to Understand Computer-Supported Collaborative Learning. *https://doi.org/10.1177/0735633121996477* 59, 7 (feb 2021), 1401–1424. https://doi.org/10.1177/0735633121996477
[114] Guzin Ozdagoglu, Gulin Zeynep Oztas, and Mehmet Cagliyangil. 2019. An application framework for mining online learning processes through event-logs. *Business process management journal* 25, 5 (2019), 860–886.
[115] Cindy Paans, Erdem Onan, Inge Molenaar, Ludo Verhoeven, and Eliane Segers. 2019. How social challenges affect children's regulation and assignment quality in hypermedia: a process mining study. *Metacognition and Learning* (2019). https://doi.org/10.1007/s11409-019-09204-9
[116] Shalini Pandey and Jaideep Srivastava. 2020. RKT: Relation-Aware Self-Attention for Knowledge Tracing. In *Proceedings of the 29th ACM International Conference on Information Knowledge Management (CIKM '20)*. Association for Computing Machinery, New York, NY, USA, 1205–1214. https://doi.org/10.1145/3340531.3411994

[117] Saurin Parikh and Hari Kalva. 2018. Predicting Learning Difficulty Based on Gaze and Pupil Response. In *Adjunct Publication of the 26th Conference on User Modeling, Adaptation and Personalization (UMAP '18)*. Association for Computing Machinery, New York, NY, USA, 131–135. https://doi.org/10.1145/3213586.3226224

[118] Nirmal Patel, Collin Sellman, and Derek Lomas. 2017. Mining frequent learning pathways from a large educational dataset. arXiv:1705.11125 https://github.com/nirmalpatel/learning_pathway_mining

[119] Robert L Peach, Sam F Greenbury, Iain G Johnston, Sophia N Yaliraki, David J Lefevre, and Mauricio Barahona. 2021. Understanding learner behaviour in online courses with Bayesian modelling and time series characterisation. *Scientific reports* 11, 1 (2021), 2823.

[120] Robert L Peach, Sophia N Yaliraki, David Lefevre, and Mauricio Barahona. 2019. Data-driven unsupervised clustering of online learner behaviour. *NPJ science of learning* 4, 1 (2019), 11–14.

[121] Sai Santosh Sasank Peri, Bodong Chen, Angela Liegey Dougall, and George Siemens. 2020. Towards Understanding the Lifespan and Spread of Ideas: Epidemiological Modeling of Participation on Twitter. In *Proceedings of the Tenth International Conference on Learning Analytics Knowledge (LAK '20)*. Association for Computing Machinery, New York, NY, USA, 197–202. https://doi.org/10.1145/3375462.3375515

[122] Eric G. Poitras, Tenzin Doleck, Lingyun Huang, Laurel Dias, and Susanne P. Lajoie. 2021. Time-driven modeling of student self-regulated learning in network-based tutors. *Interactive Learning Environments* (mar 2021), 1–22. https://doi.org/10.1080/10494820.2021.1891941

[123] Chen Qiao and Xiao Hu. 2020. A joint neural network model for combining heterogeneous user data sources: An example of at-risk student prediction. *Journal of the Association for Information Science and Technology* 71, 10 (2020), 1192–1204.

[124] Benjamin Maraza Quispe, Jhon Edwar Ninasivincha Apfata, Ricardo Carlos Qusipe Figueroa, and Manuel Alejandro Valderrama Solis. 2021. Design proposal of a personalized Dashboard to optimize teaching-learning in Virtual Learning Environments. *ACM International Conference Proceeding Series* (oct 2021), 77–84. https://doi.org/10.1145/3498765.3498777

[125] Joseph M Reilly and Chris Dede. 2019. Differences in Student Trajectories via Filtered Time Series Analysis in an Immersive Virtual World. In *Proceedings of the 9th International Conference on Learning Analytics Knowledge (LAK19)*. Association for Computing Machinery, New York, NY, USA, 130–134. https://doi.org/10.1145/3303772.3303832

[126] Jeremy Riel, Kimberly A. Lawless, and Scott W. Brown. 2018. Timing Matters: Approaches for Measuring and Visualizing Behaviours of Timing and Spacing of Work in Self-Paced Online Teacher Professional Development Courses. *Journal of Learning Analytics* 5, 1 (apr 2018). https://doi.org/10.18608/jla.2018.51.3

[127] Saman Rizvi, Bart Rienties, and Jekaterina Rogaten. 2018. Temporal Dynamics of MOOC Learning Trajectories. In *Proceedings of the First International Conference on Data Science, E-Learning and Information Systems (DATA '18)*. Association for Computing Machinery, New York, NY, USA. https://doi.org/10.1145/3279996.3280035

[128] Jairo Rodríguez-Medina, María Jesús Rodríguez-Triana, Maka Eradze, and Sara García-Sastre. 2018. Observational Scaffolding for Learning Analytics: A Methodological Proposal. In *European Conference on Technology Enhanced Learning*. Springer, Cham, 617–621. https://doi.org/10.1007/978-3-319-98572-5_58

[129] Sherry Ruan, Wei Wei, and James Landay. 2021. Variational Deep Knowledge Tracing for Language Learning. In *LAK21: 11th International Learning Analytics and Knowledge Conference (LAK21)*. Association for Computing Machinery, New York, NY, USA, 323–332. https://doi.org/10.1145/3448139.3448170

[130] John Saint, Yizhou Fan, Shaveen Singh, Dragan Gasevic, and Abelardo Pardo. 2021. Using process mining to analyse self-regulated learning: A systematic analysis of four algorithms. *ACM International Conference Proceeding Series* (2021), 333–343. https://doi.org/10.1145/3448139.3448171

[131] John Saint, Dragan Gašević, Wannisa Matcha, Nora'Ayu Ahmad Uzir, and Abelardo Pardo. 2020. Combining Analytic Methods to Unlock Sequential and Temporal Patterns of Self-Regulated Learning. In *Proceedings of the Tenth International Conference on Learning Analytics Knowledge (LAK '20)*. Association for Computing Machinery, New York, NY, USA, 402–411. https://doi.org/10.1145/3375462.3375487

[132] John Saint, Alexander Whitelock-Wainwright, Dragan Gasevic, and Abelardo Pardo. 2020. Trace-SRL: A Framework for Analysis of Microlevel Processes of Self-Regulated Learning From Trace Data. *IEEE transactions on learning technologies* 13, 4 (2020), 861–877.

[133] Mohammed Saqr and Sonsoles López-Pernas. 2021. The longitudinal trajectories of online engagement over a full program. *Computers and Education* 175 (dec 2021), 104325. https://doi.org/10.1016/J.COMPEDU.2021.104325

[134] Mohammed Saqr and Jalal Nouri. 2020. High Resolution Temporal Network Analysis to Understand and Improve Collaborative Learning. In *Proceedings of the Tenth International Conference on Learning Analytics Knowledge (LAK '20)*. Association for Computing Machinery, New York, NY, USA, 314–319. https://doi.org/10.1145/3375462.3375501

[135] Pratiti Sarkar, Kapil Kadam, and Jayesh S. Pillai. 2020. Learners' approaches, motivation and patterns of problem-solving on lines and angles in geometry using augmented reality. *Smart Learning Environments* 7, 1 (dec 2020). https://doi.org/10.1186/s40561-020-00124-9

[136] Kshitij Sharma, Zacharoula Papamitsiou, Jennifer K Olsen, and Michail Giannakos. 2020. Predicting Learners' Effortful Behaviour in Adaptive Assessment Using Multimodal Data. In *Proceedings of the Tenth International Conference on Learning Analytics Knowledge (LAK '20)*. Association for Computing Machinery, New York, NY, USA, 480–489. https://doi.org/10.1145/3375462.3375498

[137] Noa Sher, Carmel Kent, and Sheizaf Rafaeli. 2020. How 'Networked' are Online Collaborative Concept-Maps? Introducing Metrics for Quantifying and Comparing the 'Networkedness' of Collaboratively Constructed Content. *Education sciences* 10, 10 (2020), 1.

[138] Varshita Sher, Marek Hatala, and Dragan Gašević. 2019. On multi-device use: Using technological modality profiles to explain differences in students' learning. In *Proceedings of the 9th International Conference on Learning Analytics Knowledge - LAK19 (LAK19)*. ACM Press, New York, New York, USA, 1–10. https://doi.org/10.1145/3303772.3303790

[139] Yuling Shi, Zhiyong Peng, and Hongning Wang. 2017. Modeling Student Learning Styles in MOOCs. In *Proceedings of the 2017 ACM on Conference on Information and Knowledge Management (CIKM '17)*. Association for Computing Machinery, New York, NY, USA, 979–988. https://doi.org/10.1145/3132847.3132965

[140] Dongmin Shin, Yugeun Shim, Hangyeol Yu, Seewoo Lee, Byungsoo Kim, and Youngduck Choi. 2021. SAINT+: Integrating Temporal Features for EdNet Correctness Prediction. In *LAK21: 11th International Learning Analytics and Knowledge Conference (LAK21)*. Association for Computing Machinery, New York, NY, USA, 490–496. https://doi.org/10.1145/3448139.3448188

[141] Márta Sobocinski, Sanna Järvelä, Jonna Malmberg, Muhterem Dindar, Antti Isosalo, and Kai Noponen. 2020. How does monitoring set the stage for adaptive regulation or maladaptive behavior in collaborative learning? *Metacognition and Learning* 15, 2 (aug 2020), 99–127. https://doi.org/10.1007/s11409-020-09224-w

[142] Márta Sobocinski, Jonna Malmberg, and Sanna Järvelä. 2017. Exploring temporal sequences of regulatory phases and associated interactions in low- and high-challenge collaborative learning sessions. *Metacognition and Learning* 12, 2 (aug 2017), 275–294. https://doi.org/10.1007/s11409-016-9167-5

[143] Rajiv Srivastava, Girish Keshav Palshikar, Saheb Chaurasia, and Arati Dixit. 2018. What's Next? A Recommendation System for Industrial Training. *Data science and engineering* 3, 3 (2018), 232–247.

[144] Fu Rong Sun, Hong Zhen Hu, Rong Gen Wan, Xiao Fu, and Shu Jing Wu. 2019. A learning analytics approach to investigating pre-service teachers' change of concept of engagement in the flipped classroom. *Interactive Learning Environments* (2019). https://doi.org/10.1080/10494820.2019.1660996

[145] Jerry Chih Yuan Sun, Che Tsun Lin, and Chien Chou. 2018. Applying learning analytics to explore the effects of motivation on online students' reading behavioral patterns. *International Review of Research in Open and Distance Learning* 19, 2 (may 2018), 209–227. https://doi.org/10.19173/irrodl.v19i2.2853

[146] Zhong Sun, Chin Hsi Lin, Kaiyue Lv, and Jie Song. 2021. Knowledge-construction behaviors in a mobile learning environment: a lag-sequential analysis of group differences. *Educational Technology Research and Development* (apr 2021). https://doi.org/10.1007/s11423-021-09938-x

[147] Michelle Taub and Roger Azevedo. 2018. *Using Sequence Mining to Analyze Metacognitive Monitoring and Scientific Inquiry based on Levels of Efficiency and Emotions during Game-Based Learning.* Technical Report 3. 1–26 pages. https://doi.org/10.5281/ZENODO.3554711

[148] Andrew A Tawfik, Philippe J Giabbanelli, Maureen Hogan, Fortunata Msilu, Anila Gill, and Cindy S York. 2018. Effects of success v failure cases on learner-learner interaction. *Computers Education* 118 (2018), 120–132. https://doi.org/10.1016/j.compedu.2017.11.013

[149] Ahmed Tlili, Huanhuan Wang, Bojun Gao, Yihong Shi, Nian Zhiying, Chee Kit Looi, and Ronghuai Huang. 2021. Impact of cultural diversity on students' learning behavioral patterns in open and online courses: a lag sequential analysis approach. *Interactive Learning Environments* 0, 0 (2021), 1–20. https://doi.org/10.1080/10494820.2021.1946565

[150] Meng Jung Tsai and An Hsuan Wu. 2021. Visual search patterns, information selection strategies, and information anxiety for online information problem solving. *Computers and Education* 172 (oct 2021), 104236. https://doi.org/10.1016/J.COMPEDU.2021.104236

[151] Rahila Umer, Anuradha Mathrani, Teo Susnjak, and Suriadi Lim. 2019. Mining Activity Log Data to Predict Student's Outcome in a Course. In *Proceedings of the 2019 International Conference on Big Data and Education (ICBDE'19)*. Association for Computing Machinery, New York, NY, USA, 52–58. https://doi.org/10.1145/3322134.3322140

[152] Nora'ayu Ahmad Uzir, Dragan Gašević, Jelena Jovanović, Wannisa Matcha, Lisa-Angelique Lim, and Anthea Fudge. 2020. Analytics of Time Management and Learning Strategies for Effective Online Learning in Blended Environments. In *Proceedings of the Tenth International Conference on Learning Analytics Knowledge (LAK '20)*. Association for Computing Machinery, New York, NY, USA, 392–401. https://doi.org/10.1145/3375462.3375493

[153] Steven Van Goidsenhoven, Daria Bogdanova, Galina Deeva, Seppe vanden Broucke, Jochen De Weerdt, and Monique Snoeck. 2020. Predicting Student Success in a Blended Learning Environment. In *Proceedings of the Tenth International Conference on Learning Analytics Knowledge (LAK '20)*. Association for Computing Machinery, New York, NY, USA, 17–25. https://doi.org/10.1145/3375462.3375494

[154] Anouschka van Leeuwen, Nynke Bos, Heleen van Ravenswaaij, and Jurgen van Oostenrijk. 2019. The role of temporal patterns in students' behavior for predicting course performance: A comparison of two blended learning courses. *British Journal of Educational Technology* 50, 2 (2019), 921–933. https://doi.org/10.1111/bjet.12616

[155] Chenyang Wang, Weizhi Ma, Min Zhang, Chuancheng Lv, Fengyuan Wan, Huijie Lin, Taoran Tang, Yiqun Liu, and Shaoping Ma. 2021. Temporal Cross-Effects in Knowledge Tracing. In *Proceedings of the 14th ACM International Conference on Web Search and Data Mining (WSDM '21)*. Association for Computing Machinery, New York, NY, USA, 517–525. https://doi.org/10.1145/3437963.3441802

[156] Mengqian Wang, Wenge Guo, Huixiao Le, and Bo Qiao. 2020. Reply to which post? An analysis of peer reviews in a high school SPOC. *Interactive Learning Environments* 28, 5 (jul 2020), 574–585. https://doi.org/10.1080/10494820.2019.1696840

[157] Qiaosi Wang, Koustuv Saha, Eric Gregori, David Joyner, and Ashok Goel. 2021. Towards Mutual Theory of Mind in Human-AI Interaction: How Language Reflects What Students Perceive About a Virtual Teaching Assistant. In *Proceedings of the 2021 CHI Conference on Human Factors in Computing Systems (CHI '21)*. Association for Computing Machinery, New York, NY, USA. https://doi.org/10.1145/3411764.3445645

[158] Shu-Ming Wang, Huei-Tse Hou, and Sheng-Yi Wu. 2017. Analyzing the knowledge construction and cognitive patterns of blog-based instructional activities using four frequent interactive strategies (problem solving, peer assessment, role playing and peer tutoring): a preliminary study. *Educational Technology Research and Development* 65, 2 (2017), 301–323. http://www.jstor.org/stable/45018553

[159] Wei Wang, Wenxin Mu, and Juanqiong Gou. 2019. Spatial-Temporal Data Association Based Ontology Alignment Research in High Education Context. In *Proceedings of the 2019 International Conference on Big Data Engineering (BDE 2019)*. Association for Computing Machinery, New York, NY, USA, 125–130. https://doi.org/10.1145/3341620.3341640