Role of Theory in Analysing the Dynamic of Self-regulated Learning Process Based on Students' Event Logs Data: A Scoping Review

Muhammad Azani Hasibuan[1,2], Mark Reynolds[1], Sally Male[3] and Ghulam Mubashar Hassan[1]

[1] The University of Western Australia, [2] Telkom University, [3] The University of Melbourne

Students' capability to metacognitively regulate themselves—cognitively or behaviourally—in their learning plays a pivotal role in determining their academic performance. The dynamic aspect of Self-Regulated Learning (SRL) will be challenging to capture if it relies on perceptual data obtained from students through questionnaires and interviews and only at specific points in time. One potential alternative is to use an approach that captures student activity data in real-time throughout the learning period. Given the context-sensitivity involved in measuring SRL via event data, a solid theoretical foundation is essential in analysing patterns of SRL behaviour using event log data. This scoping review paper aims to identify and map how the empirical studies in this area consider SRL theory or models, not only when interpreting analytical results but also when designing instruction and interpreting SRL indicators from raw data. A thorough literature search was performed on various online databases, including Scopus, IEEExplorer, ProQuest, and Web of Science, to identify relevant studies. Following the PRISMA scoping review (PRISMA-ScR) as a protocol for the review, 39 studies published between 2012 and 2023 were included. This study found limited studies incorporating SRL theory in every analysis stage, from designing instruction to preprocessing event data and interpreting analytical models. This study also highlighted the importance of including contextual and theoretical factors when assessing self-regulatory behavioural patterns.

Keywords: self-regulated learning, event log data, educational data mining, scoping review, theory-driven analytics, learning analytics

Introduction

Self-regulated learning (SRL) is essential to students' learning and is pivotal in understanding student performance. Students who exhibit SRL can be characterised by setting goals, developing a comprehensive plan, being adaptive in implementing it, and actively self-reflecting on their learning process. SRL plays a crucial role in understanding student performance and outcome variation. It is also a dynamic process because the way students navigate their learning will vary depending on the context of the learning task and their cognitive and emotional state. Learning context can be the complexity and type of learning content, the kind of activity, and the type of learning environment. These influence the choice of SRL strategies that students will choose. Like the learning context, emotional state, motivation, and prior knowledge will also determine the form of students' SRL behaviour. In addition, the ever-changing nature of self-regulated learning implies that students' strategies can evolve and change over time. Comprehending the mechanisms underlying the dynamic aspect of SRL is essential to develop evidence-based interventions supporting students' learning processes.

The dynamic aspect of SRL will be challenging to capture if it relies on perceptual data obtained from students through self-report questionnaires and interviews, which only capture a snapshot of students' SRL (Cicchinelli et al., 2018; Siadaty, Gasevic, et al., 2016). An alternative is to use an approach that captures student activity data in real-time throughout the learning period. There are several approaches to achieve this, such as visual observation, think aloud and trace-data analysis (Winne & Perry, 2000). The capture and assessment of SRL using trace data have gained significant attention in recent years, along with the increasing adoption of online learning in educational institutions (Fan et al., 2022; Saint et al., 2021). Event log data generated by learning platforms are considered as traces representing students' learning processes. These traces can then be used as evidence for student adaptation or regulation during their learning. Due to the data-intensive nature of trace-based SRL measurement, a specific analytical strategy is required. One of the research areas that offers an analytical approach for assessing this kind of data is learning analytics. This field of research not only aims to explain or understand how learning happens but also tries to optimise or improve the learning process by giving

back the gained insight to the learner, teacher, or instruction designer. Hence, in the context of SRL measurement, the main objective of learning analytics is modelling behaviour and providing predictions and recommendations to learners to optimise their learning process.

The complexity of learning analytics when measuring SRL resides in the breadth of aspects that must be analysed in addition to data collection. This procedure is not straightforward and requires a comprehensive understanding of the underlying theoretical framework and concepts. A robust theoretical framework is crucial when considering the measurement of SRL using event data due to the inherent context sensitivity involved. The interpretation of identical empirical data can vary depending on the theoretical frameworks employed. The theoretical construct is essential for designing the learning environment, analysing the raw data, and interpreting the analysis results. The continued use of theories in all these aspects will increase the likelihood of reaching accurate interpretations of students' learning behaviour patterns (Chatti et al., 2021).

Several review studies in this field have provided valuable information regarding the measurement of SRL using learning analytics. Araka et al. (2020) contend in their literature review that there are few framework models available for measuring and facilitating SRL in the online learning environment. This framework is required to interpret digital traces left by students and draw conclusions about their learning strategies. Similar to Araka et al.(2020), Viberg et al. (2020) argue that the majority of learning analytics research aims to measure SRL behaviour, but there are few studies that use learning analytics to assist students in developing SRL skills. Saint et al. (2022) reviewed articles that investigated SRL as a temporal phenomenon using data-driven analytic models. They discovered that theoretical constructs, data collection techniques, and analytical methodologies must be considered when assessing the validity and significance of SRL studies based on analytic models. To resolve these concerns, they present a framework containing a list of queries to be considered when conducting similar research. Xu et al. (2023) discovered that SRL strategies occur in multiple phases of learning and are effective in influencing learning phases and effectively influencing students' learning performance in online and integrated learning environments, particularly in STEM disciplines. Additionally, they identified a void in the literature regarding the cognitive and affective regulation strategies of students' cognitive and affective regulation strategies in online and blended environments.

Examining the dynamic aspects of SRL through utilising learners' event data logs is an emerging area of inquiry and has the potential to be further explored. Understanding how the most recent studies employed SRL theory or models to measure the dynamic aspect of SRL will lay the groundwork for developing theory-driven learning analytics for measuring trace-based SRL. To the best of our knowledge, however, no review article describes the function of SRL theory or models in designing instruction, analysing raw data, and interpreting learning analytics models. Consequently, this study aims to investigate how SRL theory and models influence learning analytics solutions and instructional design and to emphasise the issues and future trajectories of theory-driven trace-based SRL measurement. Ultimately, advancing research on the application of data mining and learning analytics in assessing the dynamic aspect of SRL processes can have significant implications for enhancing student learning outcomes and promoting academic success.

Research questions

A conceptual framework has been devised to effectively evaluate the quality of research that employs learning analytics and data mining techniques to measure or model the dynamic nature of SRL processes, as presented in Figure 1. This conceptual framework is centred on the Self-Regulated Learning (SRL) model/theory. It highlights its crucial role as a theoretical context for modelling and assessing students' regulatory processes. In addition, the framework identifies three other concepts connected to the SRL model: the behavioural model, clickstream data (data abstraction), and instruction design. The behavioural model and SRL theory are interrelated with a reciprocal relationship. From a behavioural pattern perspective, learning analytics or data mining will provide meaningful insights if interpreted using SRL theory or models as a reference. Additionally, the identified behavioural model can be used to confirm the SRL theory/model. Furthermore, the relationship between clickstream data and the SRL theory/model indicates that event data will be meaningful if abstracted close to the SRL construct. Finally, the relationship between instruction design and SRL theory highlights the importance of SRL theory in designing learning activities and assessments to foster SRL skills. By considering the interplay between these components and the SRL model, the framework can inform the quality and validity of learning analytics when assessing the dynamics of students' regulatory processes. Based on the conceptual framework, we formulated four research questions that require elucidation. The first three questions pertain to the relationships between each dyadic aspect when measuring SRL through trace data

in online or computer-based learning environments. The final question concerns the challenges and issues in assessing SRL dynamics using event data logs.

1. RQ 1: How did previous studies incorporate the SRL model/theory into instructional design when assessing SRL using trace data in online or computer-based educational settings?
2. RQ 2: How did the past studies preprocess the raw event data logs to capture meaningful SRL events or indicators, and to what extent was the SRL model/theory and type of instructional design incorporated in that process?
3. RQ 3: How did past studies analyse students' log data to identify the SRL behaviour model or pattern? How was the SRL model/theory used when interpreting the identified behavioural models?
4. RQ 4: What are the challenges and issues when measuring SRL using learning analytics and event data? What are the gaps in the current literature?
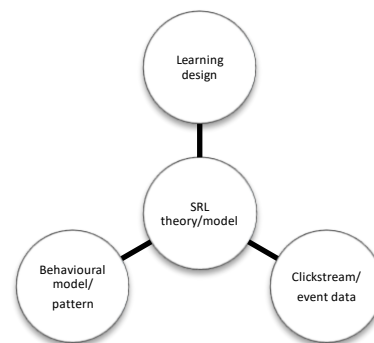
Figure 1 Conceptual model relation of SRL model/theory with learning design, event data, and behavioural model/pattern

Methodology

This study adhered to the PRISMA Scoping Review (PRISMA-ScR) protocol (Tricco et al., 2018) as a guideline for conducting a systematic review. The rationale for selecting this protocol stems from the research objective of mapping out the current frontiers of learning analytics research, focusing on assessing the dynamic aspect of self-regulated learning through the analysis of students' digital traces.

Search strategy

We implemented a comprehensive search strategy to guarantee the identification of relevant studies. We considered multiple trustworthy databases and search engines to ensure we retrieved a wide range of literature relevant to our research questions. The databases that we utilised were Scopus, Web of Science, IEEE Explorer, and ProQuest. This ensured we could access a diverse array of sources across various disciplines. To ensure the accuracy and relevance of the search, we carefully selected query terms derived from our research questions. The following expression queries were used in this study: ("learning analytics" OR "data mining" OR "educational data mining") AND "self-regulated learning" AND ("measurement" OR "measuring") AND ("online learning" OR "online course" OR "MOOC" OR "e-learning") AND ("log" OR "data logs" OR "trace data" OR "clickstream"). To further refine our search, we limited it to articles written in English and published between 2012 and 2023. This ensured we focused on the most recent and relevant literature in self-regulated learning and learning analytics.

Criteria for assessing the quality of the studies

Following the search process on multiple selected databases, we screened the articles based on their titles and abstracts to determine their eligibility. To aid in selecting and identifying relevant studies, we established a set of inclusion and exclusion criteria. The following are the inclusion criteria that we used:

1. The article should employ data mining, machine learning, or learning analytics techniques to assess or measure at least one aspect of SRL.
2. The article should use SRL as the learning construct.

3. The generated model should capture the dynamic of the student regulatory process by utilizing students clickstream data obtained from an online or computer-based learning platform.
4. The article should be an empirical study or use actual rather than simulated data.

Data extraction and analysis plan

Following the selection procedure, we extracted data from the retained articles. We created a data extraction form to extract data from the chosen publications. The form included information such as the author(s), publication year, research question(s), SRL model(s) used, analytic techniques applied, categories of data collected, and results about the dynamic aspect of SRL. Using narrative synthesis, the extracted data and quality assessment results were synthesised and analyzed. This strategy entailed identifying patterns, themes, and relationships across the selected articles and providing a detailed description of the most important findings. In addition, we utilised a tabular summary to present the key characteristics of the studies, including the research questions, SRL model(s) or theory(s) used, analytic techniques employed, categories of data collected, and results about the dynamic aspect of SRL.

Results

Study selection

Following the PRISMA-ScR (), we identified 391 articles from various databases and search engines and 23 articles from other methods (citation searching) in the initial identification phase. In total, we identified 414 articles. Removing duplicate articles resulted in a final selection of 391 unique articles. Subsequently, a rigorous screening was conducted to identify articles satisfying the predetermined inclusion criteria. After implementing the inclusion criteria, 39 articles (27 journal articles and 12 conference articles) were retained.
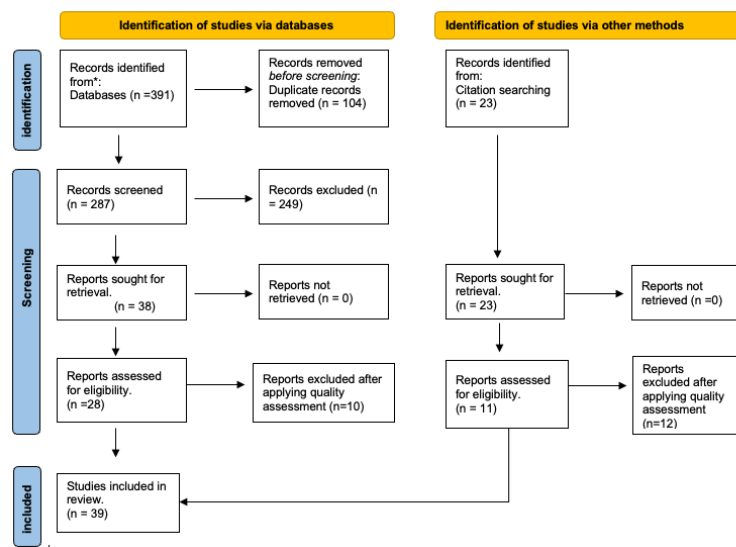

Figure 2 Flow of Selection process using PRISMA-ScR

Characteristics of studies

An examination of several studies on self-regulated learning (SRL) analysis in online and computer-based learning reveals several key characteristics in terms of period, location, educational level, course categories, and study design. Our analysis indicates that most studies (N=35) were conducted in the most recent six-year period between 2018 and 2023, with only a small number (N=4) being conducted from 2012 to 2017. Regarding the study's location, 35 studies mentioned the location, while only four did not. The highest number of studies (N=8) were conducted in Australia, followed by the United States (N=7), Canada (N=4), Chile, China, Germany, the Netherlands, and Taiwan (N=2 each). The remainder were held in Malaysia, Spain, the United Kingdom, and Switzerland. Regarding educational level, most studies (N= 33) were conducted in higher education institutions, indicating that researchers have focused on SRL conduct in this context. However, two

studies were carried out in a workplace setting and four in K-12 education, which suggests that researchers are also interested in understanding SRL in other settings. In terms of course categories, most studies took place in STEM-related units (N=25). Additionally, our analysis shows that authentic studies accounted for the majority (N=28) of the research, while 11 studies were conducted in experimental contexts. Experimental studies suggest that researchers are interested in testing the effectiveness of interventions or approaches that enhance SRL. In contrast, the use of real-world learning contexts suggests that researchers are interested in exploring SRL in authentic settings. Overall, these findings provide insights into the characteristics of studies that have examined the use of data mining and learning analytics to identify or forecast SRL conduct in online education.

In addition to examining the characteristics of the studies, we analysed the type of their instructional context, including the type of learning platform, the type of learning strategy and the type of instructional approach. Regarding the learning platform type, most studies (N=17) employed Learning Management Systems (LMS). This was followed by Computer-based learning Platforms (N=8), MOOC platforms (N=6), Intelligent Tutoring Systems (ITS) (N=3), and various other platforms such as Game-based learning platforms, open-ended learning, online learning journals, and online collaborative learning. Regarding the type of learning strategy, most studies (N=32) employed an instruction-based learning approach, wherein students studied independently after receiving instructions in the form of text or multimedia content. Four out of thirty-two studies employed a blended learning strategy in which students also learned face-to-face in class in addition to independent learning. Problem-based learning (N=6) occupies the second most common strategy, in which students are given a set of problems and then asked to solve them independently or in groups. The remaining studies employed apprenticeship-based learning. In terms of instructional approach, most studies (N=38) were conducted in the context of individual learning, while only one study involved collaborative learning.

RQ 1: Role of SRL theory in designing learning environment

Our examination of numerous studies revealed that most instructional design and learning platforms (N=31) did not intentionally incorporate self-regulated learning theories or models into their course instruction, and only eight studies (Azevedo & Kinnebrew, 2012; Kinnebrew et al., 2017; Ng et al., 2023; Segedy et al., 2015; Siadaty, Gašević, et al., 2016; Sun et al., 2023; Taub et al., 2018; Wong et al., 2021) specifically designed their learning platform to support SRL behaviour in students. Within this subset of studies, four out of eight studies used Winne et al. (1998) SRL model as a reference for designing the instructions. Three studies utilised Zimmerman's (1989) models, while only one study employed Roscoe et al. (2020) SRL model as a reference. Studies incorporating SRL theories into their instructional design employ various strategies to encourage SRL behaviour. For instance, (Wong et al., 2021) designed a tool to interrupt students' study sessions with SRL prompts. These prompts are used to activate learning strategies through indirect instruction. These prompts will encourage students to reflect on their learning process and suggest necessary SRL activities (Wong et al., 2021). Another approach is to define learning tasks based on the SRL process. For instance, Kinnebrew et al. (2017) designed the Betty's Brain platform to train students to impart a concept to a virtual agent (named Betty) via the construction of a concept map. They characterised each cognitive task in Betty's brain as an instance of SRL processes. Other studies intentionally designed a learning platform to foster SRL behaviour in students. For example, Taub et al. (2018) used Cristal Island, a game-based learning platform that aimed to support students' SRL behaviour.

RQ 2: Role of SRL theory in data preprocessing

To answer the second research question, which focuses on the association between the characteristics of student traces and learning analytics solutions in assessing the dynamic aspect of students' SRL in online learning, we thoroughly examined all the information related to the data analyses in the reviewed studies.
We carefully examined each study to find the pre-processing and analysis procedure information. Then, we classified the data source based on the type of data, pre-processing, and feature or attribute type. We also classified each study using seven different labels representing the data source type to differentiate the data type. It is important to note that each study could have more than one data source. After analysing the data sources used in the studies, we found that more than half of the studies used students' activity logs and data from other sources. Student grades were the most frequent additional data source, with nine studies using them, followed by survey responses (N=6) and demographic data (N=4).

Moreover, most studies performed several pre-processing activities on the data sources. Studies using event log data perform data filtering, abstraction, and learning section identification. Regarding the event data abstraction

process, we discovered that SRL theory was applied in most studies (N=23). These studies utilised SRL theory to interpret or label students' actions recorded in an event data log. These studies used various SRL dimensions such as aspects, processes, and strategies to label students' actions or series of actions. Most studies used SRL processes (macro processes) such as orientation, planning, executing, and evaluating as the labels. Some studies (Fan et al., 2022; Fan, Saint, et al., 2021; Lim et al., 2023; Quick et al., 2023; Saint et al., 2020; Siadaty, Gašević, et al., 2016; Sun et al., 2023; Wang et al., 2023) further decomposed each SRL macro process into several micro-processes, such as goal setting, making personal plans, working on the task, and reflection. We also identified two common approaches to event data abstraction: top-down and button-up. The top-down approach, also known as semi-supervised labelling, uses predefined libraries to determine the type of micro process or process from a set of event data. For example, the study by Saint et al. (2020) and Siadaty, Gasevic, et al. (2016). The second approach is bottom-up or unsupervised learning, where data events are grouped based on the similarity of sequence patterns. These pattern groups are then interpreted as learning tactics, the smallest instantiation of an SRL process. This type of approach is used by Fan, Saint, et al. (2021 and Matcha et al. (2020). Several of these studies (N=12) also incorporated frequency-based features as a supplement to sequence data. These studies tended to calculate the frequency and duration of each activity carried out by students. To conduct effective trace data analysis, it is necessary to establish a precise definition of a learning session or episode. A learning session or episode is a discrete unit that depicts a student's learning activity in this context. It consists of a series of learning activities within a specific time frame. The definitions of learning sessions in the reviewed studies vary, such as biweekly (N=1), 24 hours (N=1), duration of learning module (N=3), as well as sessions delimited by login and logout (N=2), and sessions separated by idle activities (ranging between 20 to 40 minutes) (N=14).

RQ 3: Role of SRL theory in exploring produced behavioural models/patterns

To explore how SRL theory or models are used to interpret learning analytics outcomes, we extracted information related to learning analytics solutions, including the analytical objective, data analysis procedures, the role of SRL in outcome interpretation, and the algorithm's name. Our analysis further revealed that most studies (N=23) used SRL models when interpreting the behavioural models or patterns. Specifically, more than a third of the studies (N=13) used SRL models to interpret the pattern of students' actions to identify students' SRL strategies or tactics. In addition, we discovered several ways to portray SRL behaviour through event log data modelling. SRL behaviour pattern is the most common type used as the SRL model representation (N=25), followed by learning strategies (N=14). Out of the studies that utilised SRL patterns, 10 analysed the behavioural patterns of a process model, while the remaining studies depicted a sequence pattern of SRL activities. Several studies utilised learning strategies (Fan, Matcha, et al., 2021; Matcha et al., 2019; Matcha, Gašević, et al., 2020) as an indication of the dynamic SRL behaviour of the students. Students create a pattern of event sequences, which is the basis for the learning technique. After that, the various learning strategies utilised by the students are isolated by clustering their tactics into several categories. In the meantime, the SRL process model is used to characterise the transitions between different processes or micro-processes. To build this process model, process or sequential mining techniques were utilised. This process model is frequently utilised in research that compares the SRL processes of two or more groups of students or that takes place between different learning sessions.

## RQ4: Challenges in measuring the dynamic of SRL using trace data

To analyse the issues and challenges in the reviewed studies, we compiled a table of every problem and challenge the authors reported in their articles. We then clustered these topics based on common attributes to identify the major themes. Based on our observations, we found that researchers commonly reported two themes of issues in their studies: the subjectivity of analysis and the lack of variety in data sources.The subjectivity of analysis is related to the interpretation of cluster patterns of students' actions. Identified clusters or groups of students' actions can be interpreted differently depending on the purpose of the analysis, and these clusters may differ when replicated in different learning contexts. This is why identified clusters cannot be generalised, and researchers need to be cautious when interpreting their findings. The second common issue is the lack of variety in data sources. Although students' trace data were the primary data source in the reviewed studies, they have limitations in capturing the entire aspects of SRL, especially unobservable aspects such as motivation and emotion. Moreover, assessing SRL group behaviour is challenging when only relying on students' digital traces.

Regarding future directions, most studies recommended using multisource data to enrich the measurement of SRL. Future studies should not limit their data to traces that only come from the learning management system,

as in authentic learning settings, a course could have and use more than one learning platform with unique purposes, and incorporate data associated with affective or motivational aspects of students' learning, which could be captured using wearable sensors worn by students during their learning. Using data from various sources could provide new insights into understanding students' SRL process. Another common direction recommended by the reviewed studies is to formalise the SRL process model. This direction was recommended by studies that used process mining as their analytical tool. Most process mining studies were exploratory, focusing on identifying SRL process model patterns. However, a formalisation of the SRL process model is required to conduct confirmatory studies. In confirmatory studies of process mining, a formalised SRL process model could be used as a reference model to identify deviations in actual students' SRL processes. This formalised model could also be used as a learner model in an intelligent tutoring system (ITS) that would trigger real-time and personalised feedback to students when they deviate from the model. In summary, the reviewed studies suggested that future research should explore multisource data and formalise the SRL process model to enhance the accuracy and applicability of SRL analysis.

Discussion

This study maps the role of SRL theories or models in the instruction design, data pre-processing, and model analysis aspects when measuring SRL through trace data in online or computer-based learning environments. We divided each study into three components: the learning environment aspect, which is the context of the learning activity. The data preprocessing aspect, where SRL indicators are defined, and the model or pattern aspect, which results from the process analysis. Using these aspects as a reference, we found thirteen studies incorporated SRL theories or models in all aspects of analytics, as shown in Figure 3. While the remainder of the studies partially utilised the SRL theory. More than half of the studies (N=23) interpreted event data using SRL theory. Eighteen of these subsets employ learning environments in which the SRL theory is not explicitly applied, while only five studies in this subset employ instructional design based on the SRL theory. Concerning the use of SRL theory in data analytics, twenty-three studies interpreted the behavioural model or pattern using SRL theory. However, only ten studies constructed the model using SRL-based event abstraction.
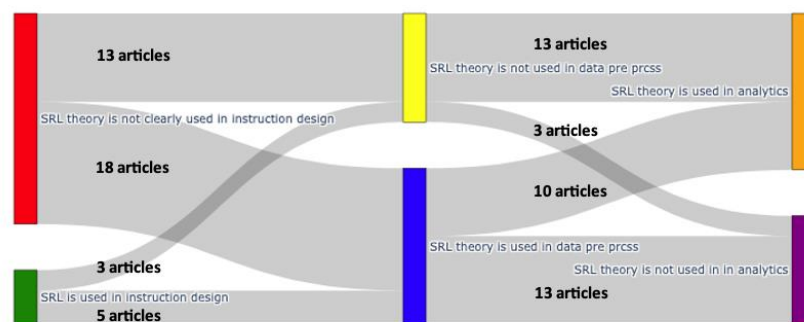
Figure 3 The relationship of the role of SRL theory in instruction design, data pre-processing, and analytical model or pattern

Although SRL has multiple categories of constructs on a theoretical level, most of the studies we reviewed focused on the cognitive or meta-cognitive aspects of SRL analysis. When analysing other SRL constructs, such as emotion and motivation, these studies drew on information from other sources. The study by Taub et al. (2018), for instance, utilised EDA (Electrodermal Activity) data as a data source for measuring the affective state of students. When analysing the motivational construct of students, one study (Wong et al., 2021) employed questionnaire data. This finding indicates that the event data log cannot capture the complete SRL behaviour (cognitive, emotional, and motivational). Multi-modal analytics, a method that utilises multiple data sources, is a prospective method for comprehensively analysing SRL (Azevedo & Gasevic, 2019; Chango et al., 2022; Hanna Järvenoja et al., 2020). In addition to having several aspects that are regulated, and although there are several SRL theories related to the phases or stages of regulation, all theories can be grouped into three main phases: preparation, engagement, and appraisal. Based on these three phases, we found that all studies analysed SRL behaviour in the performance phase, 26 studies analysed the preparation phase, and 24 studies analysed the appraisal phase. So overall, there were few studies (N = 26) using all phases in analysing SRL behaviour.

Almost all the studies we reviewed analysed SRL behaviour at the individual level. Theoretically, as stated by Hadwin et al. (2013), SRL behaviour consists of 3 levels: the most basic level is the individual level, the level

above that is co-SRL, and the third level is shared SRL. The second and third levels are usually found in collaborative learning contexts. Several articles we reviewed had a collaborative learning context but only analysed SRL behaviour at the individual level. Regarding the data pre-processing phase, most of the studies we reviewed used SRL as a lens to interpret the raw data from the event logs. This process is also known as the event abstraction process. We identified two approaches to interpreting the raw data: top-down (supervised abstraction) and bottom-up (unsupervised abstraction). In supervised abstraction, the event or sequence of events is labelled using a predefined label library. The type of label commonly used in this approach is the SRL process. The second approach is to interpret the data in a bottom-up approach. This approach uses data mining or machine learning techniques to cluster similar event patterns. These identified groups are then labelled using SRL theory as justification in the form of learning tactics. The identification of learning tactics in clustering data results can be biased due to researchers' subjective interpretation. The tactics identified may differ from study to study because they are closely related to the nature and content of a course. For instance, the tactics used in a course dominated by videos and assessments (as seen in Matcha, Gasevic, et al., 2020; Zhang et al., 2022; Zheng et al., 2019) will differ from those used in a course dominated by problem-solving activities (as seen in Kinnebrew et al., 2017; Wang et al., 2023). This suggests that it is difficult to generalise the identified learning tactics due to variations in the learning environment. When it comes to analysing the dynamic SRL behaviour, there are different approaches we identified. Studies that employed learning strategies viewed the shift between learning strategies as a representation of the SRL behaviour dynamics (Fan, Matcha, et al., 2021; Taub et al., 2018). Meanwhile, studies that used process models analysed the dynamics of process models by superimposing with high-performing students or theory (Cerezo et al., 2020). This approach is known as confirmatory process analysis (van der Aalst, 2016).

Conclusion

In conclusion, this paper has reviewed the current state of assessing the dynamic aspect of SRL based on students' traces stored in the event data log. We discovered that the SRL theory is utilised in various stages of analysis - from designing instructions preparing data, to interpreting models. However, only a few studies have created a comprehensive framework for measuring the SRL process. Some studies analyse dynamic SRL behaviour by focusing on shifts between strategies. Those that use SRL process models employ a confirmatory process analysis approach. All studies concentrate on analysing SRL at the individual level, with only one study examining SRL at the group level. We also noted that the bottom-up analytical approach, which uses a clustering algorithm, tends to produce a model that is challenging to generalise. This was observed in studies that utilised learning tactics and strategies as a manifestation of SRL behaviour. Additionally, most studies only use event data logs to address cognitive or metacognitive SRL and must incorporate different data sources to analyse SRL's emotional or motivational aspect.

Based on these findings, we argue that theoretical and instructional context plays a crucial role in assessing or capturing the dynamics of cognitive and metacognitive aspects of SRL left by students in their digital traces. Specifically, the theoretical and instructional context could augment fine-grained event data logs, making meaningful high-level activities easily identified. These identified meaningful activities will produce a meaningful model that explains the dynamic aspect of SRL. However, the existing studies still lack systematic guidelines on augmenting event log data that consider both theoretical and instructional contexts, especially in collaborative learning. Furthermore, a limited metric quantifies the quality of the collaboration process that can be inferred from the event data log. Hence, future studies should focus on establishing a formal and systematic approach to augmenting event data logs by incorporating theoretical and instructional context. Future studies could also analyse the SRL group dynamic using innovative metrics that could be extracted from multi-modal event data logs.

Although this review article is comprehensive, we would like to highlight a few limitations. Firstly, the keywords we used may not encompass all relevant articles related to our research topic. Some studies may use different terms but still cover the same concepts we explored. Secondly, although we have employed a rigorous approach in our coding process, we realise a potential bias when interpreting the studies. Despite these limitations, this review article provides insight into the current landscape of data mining use in SRL research, which is relevant for understanding self-regulated learning behaviour in online learning. The emerging sequence and process mining approaches were highlighted as potential areas for future exploration in the measurement of SRL dynamics.

References

Araka, E., Maina, E., Gitonga, R., & Oboko, R. (2020). Research trends in measurement and intervention tools for self-regulated learning for e-learning environments—Systematic review (2008–2018). Research and Practice in Technology Enhanced Learning, 15(1), 6. https://doi.org/10.1186/s41039-020-00129-5

Azevedo, R., & Gasevic, D. (2019). Analyzing Multimodal Multichannel Data about Self-Regulated Learning with Advanced Learning Technologies: Issues and Challenges. COMPUTERS IN HUMAN BEHAVIOR, 96, 207–210. https://doi.org/10.1016/j.chb.2019.03.025

Azevedo, R., & Kinnebrew, J. S. (2012). Identifying students' characteristic learning behaviors in an intelligent tutoring system fostering self-regulated learning. Proceedings of the 5th International Conference on Educational Data Mining, EDM 2012. Scopus. https://www.scopus.com/inward/record.uri?eid=2-s2.0-85084015233&partnerID=40&md5=8f979e5571c5483e1a432ebfa33c5373

Cerezo, R., Bogarín, A., Esteban, M., & Romero, C. (2020). Process mining for self-regulated learning assessment in e-learning. Journal of Computing in Higher Education, 32(1), 74–88. Scopus. https://doi.org/10.1007/s12528-019-09225-y

Chango, W., Lara, J. A., Cerezo, R., & Romero, C. (2022). A review on data fusion in multimodal learning analytics and educational data mining. WIREs Data Mining and Knowledge Discovery, 12(4). https://doi.org/10.1002/widm.1458

Chatti, M. A., Volkan Yücepur, Muslim, A., Guesmi, M., & Shoeb Joarder. (2021). Designing Theory-Driven Analytics-Enhanced Self-Regulated Learning Applications. Advances in Analytics for Learning and Teaching, 47–68. https://doi.org/10.1007/978-3-030-81222-5_3

Cicchinelli, A., Veas, E., Pardo, A., Pammer-Schindler, V., Fessl, A., Barreiros, C., & Lindstädt, S. (2018). Finding traces of self-regulated learning in activity streams. Proceedings of the 8th International Conference on Learning Analytics and Knowledge, 191–200. https://doi.org/10.1145/3170358.3170381

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

Fan, Y., Matcha, W., Uzir, N. A., Wang, Q., & Gasevic, D. (2021). Learning Analytics to Reveal Links Between Learning Design and Self-Regulated Learning. International Journal of Artificial Intelligence in Education, 1–42. https://doi.org/10.1007/s40593-021-00249-z

Fan, Y., Saint, J., Singh, S., Jovanovic, J., & Gašević, D. (2021). A learning analytic approach to unveiling self-regulatory processes in learning tactics. Null. https://doi.org/10.1145/3448139.3448211

Hadwin, A. F., Järvelä, S., & Miller, M. (2013). Self-Regulated, Co-Regulated, and Socially Shared Regulation of Learning. In Handbook of Self-Regulation of Learning and Performance. Routledge. https://doi.org/10.4324/9780203839010.ch5

Hanna Järvenoja, Järvenoja, H., Malmberg, J., Törmänen, T., Mänty, K., Haataja, E., Ahola, S., & Järvelä, S. (2020). A Collaborative Learning Design for Promoting and Analyzing Adaptive Motivation and Emotion Regulation in the Science Classroom. Frontiers in Education, 5. https://doi.org/10.3389/feduc.2020.00111

Higgins, N. L., Rathner, J. A., & Frankland, S. (2021). Development of self-regulated learning: A longitudinal study on academic performance in undergraduate science. Research in Science & Technological Education, 1–25. https://doi.org/10.1080/02635143.2021.1997978

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

Lim, L., Bannert, M., van der Graaf, J., Singh, S., Fan, Y., Surendrannair, S., Rakovic, M., Molenaar, I., Moore, J., & Gašević, D. (2023). Effects of real-time analytics-based personalized scaffolds on students' self-regulated learning. Computers in Human Behavior, 139. Scopus. https://doi.org/10.1016/j.chb.2022.107547

Lucieer, S. M., Jonker, L., Visscher, C., Rikers, R. M. J. P., & Themmen, A. P. N. (2016). Self-regulated learning and academic performance in medical education. Medical Teacher, 38(6), 585–593. https://doi.org/10.3109/0142159X.2015.1073240

Matcha, W., Gašević, D., Jovanović, J., Uzir, N. A., Oliver, C. W., Murray, A., & Gasevic, D. (2020). Analytics of learning strategies: The association with the personality traits. 151–160. Scopus. https://doi.org/10.1145/3375462.3375534

Matcha, W., Gasevic, D., Uzir, N. A., Jovanovic, J., & Pardo, A. (2019). Analytics of learning strategies: Associations with academic performance and feedback. 461–470. https://doi.org/10.1145/3303772.3303787

Matcha, W., Gasevic, D., Uzir, N. A., Jovanovic, J., Pardo, A., Lim, L., Maldonado-Mahauad, J., Gentili, S., Perez-Sanagustin, M., & Tsai, Y.-S. (2020). Analytics of learning strategies: Role of course design and

delivery modality. JOURNAL OF LEARNING ANALYTICS, 7(2), 45–71. https://doi.org/10.18608/jla.2020.72.3

Ng, J. T. D., Liu, Y., Chui, D. S. Y., Man, J. C. H., & Hu, X. (2023). Leveraging LMS logs to analyze self-regulated learning behaviors in a maker-based course. LAK23: 13th International Learning Analytics and Knowledge Conference, 670–676. https://doi.org/10.1145/3576050.3576111

Quick, J. D., Motz, B., & Morrone, A. (2023). Lost in translation: Determining the generalizability of temporal models across course contexts. LAK23: 13th International Learning Analytics and Knowledge Conference, 273–283. https://doi.org/10.1145/3576050.3576092

Roscoe, R. D., McNicol, S. A., Bhat, K. R., & Craig, S. D. (2020). A heuristic evaluative framework for self-regulated learning design. Null. https://doi.org/10.1177/1071181320641429

Saint, J., Alexander Whitelock-Wainwright, Whitelock-Wainwright, A., Gasevic, D., Pardo, A., & Pardo, A. (2020). 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. https://doi.org/10.1109/tlt.2020.3027496

Saint, J., Fan, Y., Gašević, D., & Pardo, A. (2022). Temporally-focused analytics of self-regulated learning: A systematic review of literature. Computers and Education: Artificial Intelligence, 3, 100060. https://doi.org/10.1016/j.caeai.2022.100060

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

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

Siadaty, M., Gasevic, D., & Hatala, M. (2016). Trace-based Micro-analytic Measurement of Self-Regulated Learning Processes. Journal of Learning Analytics, 3(1). https://doi.org/10.18608/jla.2016.31.11

Sun, J. C.-Y., Liu, Y., Lin, X., & Hu, X. (2023). Temporal learning analytics to explore traces of self-regulated learning behaviors and their associations with learning performance, cognitive load, and student engagement in an asynchronous online course. Frontiers in Psychology, 13, 1096337. https://doi.org/10.3389/fpsyg.2022.1096337

Taub, M., Azevedo, R., Bradbury, A. E., Millar, G. C., & Lester, J. (2018). Using sequence mining to reveal the efficiency in scientific reasoning during STEM learning with a game-based learning environment. Learning and Instruction, 54, 93–103. Scopus. https://doi.org/10.1016/j.learninstruc.2017.08.005

Tricco, A. C., Lillie, E., Zarin, W., O'Brien, K. K., Colquhoun, H., Levac, D., Moher, D., Peters, M. D. J., Horsley, T., Weeks, L., Hempel, S., Akl, E. A., Chang, C., McGowan, J., Stewart, L., Hartling, L., Aldcroft, A., Wilson, M. G., Garritty, C., … Straus, S. E. (2018). PRISMA Extension for Scoping Reviews (PRISMA-ScR): Checklist and Explanation. Annals of Internal Medicine, 169(7), 467–473. https://doi.org/10.7326/M18-0850

Van der Aalst, W. (2016). Process Mining. Springer Berlin Heidelberg. https://doi.org/10.1007/978-3-662-49851-4

Viberg, O., Khalil, M., & Baars, M. (2020). Self-regulated learning and learning analytics in online learning environments: A review of empirical research. Null. https://doi.org/10.1145/3375462.3375483

Wang, T., Li, S., Huang, X., & Lajoie, S. P. (2023). Task complexity affects temporal characteristics of self-regulated learning behaviours in an intelligent tutoring system. Educational Technology Research and Development. https://doi.org/10.1007/s11423-023-10222-3

Winne, P. H., Hadwin, A. F., & Hadwin, A. E. (1998). Studying as self-regulated learning. In Metacognition in Educational Theory and Practice.

Winne, P. H., & Perry, N. E. (2000). Measuring Self-Regulated Learning. 531–566. https://doi.org/10.1016/b978-012109890-2/50045-7

Wong, J., Baars, M., de Koning, B. B., & Paas, F. (2021). Examining the use of prompts to facilitate self-regulated learning in massive open online courses. Computers in Human Behavior, 115, 106596. https://doi.org/10.1016/j.chb.2020.106596

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

Xu, Z., Zhao, Y., Zhang, B., Liew, J., & Kogut, A. (2022). A meta-analysis of the efficacy of self-regulated learning interventions on academic achievement in online and blended environments in K-12 and higher education. Behaviour & Information Technology, 1–21. https://doi.org/10.1080/0144929X.2022.2151935

Zhang, T., Taub, M., Chen, Z., & Assoc Comp Machinery. (2022). A multi-level trace clustering analysis scheme for measuring students' self-regulated learning behavior in a mastery-based online learning environment (WOS:000883327600019). 197–207. https://doi.org/10.1145/3506860.3506887

Zheng, J., Xing, W., & Zhu, G. (2019). Examining sequential patterns of self- and socially shared regulation of STEM learning in a CSCL environment. Computers & Education, 136, 34–48. https://doi.org/10.1016/j.compedu.2019.03.005

Zimmerman, B. J. (1989). Models of Self-Regulated Learning and Academic Achievement. In B. J. Zimmerman & D. H. Schunk (Eds.), Self-Regulated Learning and Academic Achievement (pp. 1–25). Springer New York. https://doi.org/10.1007/978-1-4612-3618-4_1