Towards a Methodological Framework for Sequence Analysis in the Field of Self-Regulated Learning

Stijn Van Laer[a] & Jan Elen[a]

[a]KU Leuven, Belgium

Abstract

In recent decades, conceptualizations and operationalizations of self-regulated learning (SRL) have shifted from SRL as an aptitude to SRL as an event. Alongside this shift, increased technological capability has introduced computer log files to the investigation of SRL, uncovering new research avenues. One such avenue investigates the time-related characteristics of SRL through learners' behavioural sequences. Although sequence analysis is still relatively new in SRL research, other fields have fruitful traditions in its application and may serve as a basis for applications in the field of SRL. Ten years of investigating SRL through sequence analysis have produced a wide range of methodological approaches. While this variety of methods illustrates the diversity of opportunities, it also indicates the lack of consensus regarding the most appropriate approaches often resulting in difficult to understand methods and non-transparent ways of reporting. Since the introduction of sequence analysis in the field of SRL, researchers have been emphasizing the need for a methodological framework to guide its application. Yet, to date, no such framework has been proposed, hindering our progress through (1) transparent methods and (2) comparative studies to (3) empirical and ecological applications. To help overcome this issue, this manuscript discusses the basis of a methodological framework for the use of sequence analysis in SRL research. We first make a case for why such a framework is necessary; secondly, we propose a set of guidelines which could serve as a starting point for the construction of a framework.

Keywords: Computer log files; Sequence analysis; Self-regulated learning; Methodological framework

Corresponding author: Stijn Van Laer, Centre for Instructional Psychology and Technology, KU Leuven, Dekenstraat 2, bus 3773, B-3000 Leuven, Belgium, stijn.vanlaer@kuleuven.be DOI: https://doi.org/10.14786/flr.v6i3.367

Acknowledgments

1. Introduction

Over the last five decades, multiple theoretical conceptualizations and practical operationalizations have been proposed for self-regulated learning (SRL), shifting the focus from SRL as an aptitude to SRL as an event (e.g., Endedijk, Brekelmans, Sleegers, & Vermunt, 2016; Panadero, Klug, & Järvelä, 2016; Winne, 2016). Besides this shift, technological developments have meant that computer log files now have a role to play in investigations of learners' SRL. From both theoretical and practical perspectives, computer log files are an interesting avenue for investigating learners' SRL (e.g., Azevedo & Hadwin, 2005; Winne, 2005; Zimmerman & Schunk, 2001). On the one hand, their sequenced structure means that computer log files possess time-related characteristics relevant to the current conceptualization of SRL as an event (e.g., Azevedo, 2014; Ben-Eliyahu & Bernacki, 2015; Molenaar & Järvelä, 2014). On the other hand, their unobtrusive nature enables us to observe traces of SRL in learners' behaviour in ecologically valid contexts (e.g., Bourbonnais et al., 2006; Hine, 2011).

While sequence-based analysis has only become popular as a means of investigating the time-related characteristics of SRL within the last ten years, other fields of research (e.g., bioinformatics, chemistry, marketing, and sociology) have longstanding traditions in the use of such analyses. Insights gained from these fields may serve as a basis for applying sequence analysis in investigations of SRL (e.g., Perer & Wang, 2014; Winne & Baker, 2013). A decade of log file sequence analysis in SRL research has produced a large amount of relevant work (e.g., Azevedo, Taub, & Mudrick, 2015; Bannert, Molenaar, Azevedo, Järvelä, & Gašević, 2017; Molenaar & Järvelä, 2014; Roll & Winne, 2015) and a variety of methodological approaches. While theory-driven approaches for example prefer to recode log files to theoretically meaningful events, predefine the length of an ideal sequence, or set the threshold for significance (e.g., Roll & Winne, 2015; Winne, 2010), data-driven approaches often prefer to extract the most common sequences from the data, regardless of their content and length (e.g., Bannert et al., 2017; Beheshitha, Gašević, & Hatala, 2015). Differences with regard to the statistical analyses used can also be found. Some researchers investigate the occurrence of for example particular sub-sequences as varying from learner to learner and apply multi-level analysis (e.g., Taub, Azevedo, Bouchet, & Khosravifar, 2014; Taub, Azevedo, Bradbury, Millar, & Lester, 2017),while others focus on clusters of learners and instead apply chi-square analysis (e.g., Van Laer & Elen, 2016; Van Laer, Jiang, & Elen, 2018) or variance analysis. Still others argue that statistical analysis based on sub-sequences is insufficient to establish a full picture of learners' learning patterns and prefer to use stochastic models based on the entire sequences to operationalize the investigation of learners' behaviour (e.g., Bannert, Sonnenberg, Mengelkamp, & Pieger, 2015; Sonnenberg & Bannert, 2015). This multitude of approaches demonstrates not only the diversity of opportunities, but also the lack of consensus regarding the most appropriate methods. This lack of consensus often results in fragmentation, leading to non-transparent research practices and research reports, hampering the validation and testing of methods and thus the advancement of the investigation of SRL through sequence analysis. In line with this observation, researchers have been emphasizing the need for a methodological framework to guide the application of log file sequence analysis in SRL research since 2014 (e.g., Azevedo, 2014; Bannert, Reimann, & Sonnenberg, 2014; Molenaar & Järvelä, 2014). Such a methodological framework could, on the one hand, provide a decision-tree-like approach to choosing which analysis to perform when (Schnaubert, Heimbuch, & Bodemer, 2016) and, on the other hand, offer guidelines for reporting on each of the steps taken and considerations made. Yet, to date, no methodological frameworks have been proposed (e.g., Segedy & Biswas, 2015; Winne, 2014), hindering our ability to validate, duplicate, and so to demonstrate progress in the use of sequence analysis in SRL research and our search for the most appropriate methods (Kuhn, 2012).

Therefore, in this manuscript we discuss a methodological framework for the application of sequence analysis in the field of SRL. To do so, we first make a case for why such a methodological framework is

necessary. Secondly, we propose a set of guidelines which may serve as a starting point for the construction of a framework. With a methodological framework in place, the investigation of time-related characteristics in SRL using sequence analysis could evolve towards (1) the use of transparent methods, (2) comparative studies, and (3) empirical and ecological applications, supporting both research and practice. In what follows, we first define sequence analysis, elaborate on its link to SRL and introduce the most common phases in its operationalization, providing an illustrative example from one of our own studies. The illustrative example used in this manuscript is not intended as a good practice, but a demonstration of the complexity of sequence analyses and the decisions to be made. At the end of this introductory section, we outline the operational efforts made in the search for tangible proof of progress in sequences analysis for the investigation of learners' SRL as a method. Based on insights gathered from the introductory section, the second section proposes a set of guidelines upon which framework construction can be based. In the third and final section, we elaborate on the implications for research and practice and suggest further directions in the construction of a methodological framework for sequence analysis in the field of SRL.

1.1 Sequence analysis

A sequence (β) is an ordered list of elements (β = < A, C, B, D, E, G, C, E, D, B, G >) (Zhou, Xu, Nesbit, & Winne, 2010). Such elements can be physical, behavioural, or conceptual in nature. The analysis of a sequence makes it possible to discover hidden time-related relations between different sequences, parts of these sequences, and the individual elements within these sequences (Antunes & Oliveira, 2001). Sequence analysis therefore is indispensable in many application domains (e.g., bioinformatics, chemistry, marketing, sociology, and education) (Liu, Dev, Dontcheva, & Hoffman, 2016) and approaches are plentiful. For example in bioinformatics sequence analysis is the process of investigating a deoxyribonucleic acid (DNA) sequence to understand its features, function, structure, or evolution (e.g., Lubahn et al., 1988; Stackebrandt & Goebel, 1994). In chemistry, sequence analysis comprises the determination of the sequence of a polymer formed of several monomers (e.g., Martin, Shabanowitz, Hunt, & Marto, 2000; Van Krevelen & Te Nijenhuis, 2009). In marketing, sequence analysis on its turn is often used in analytical customer relationship management applications, such as next product to buy models (e.g., Kumar, Venkatesan, & Reinartz, 2004; Prinzie & Van den Poel, 2007). In sociology, sequence methods are increasingly used to study life-course and career trajectories, patterns of organizational and national development, conversation and interaction structure, and the problem of work and family synchrony (e.g., Bonin, Vogel, & Campbell, 2014; Stark & Vedres, 2012). Finally, in recent years the field of education also gained interest in the investigation of sequence data. Methods have been increasingly used in the context of data analysis to investigate learning processes (Reimann, Markauskaite, & Bannert, 2014). One distinct area of learning research in which sequence-analysis methods have been used is SRL-research, in particular for studying regulation and metacognition in students' learning through computer log files (e.g., Azevedo, Moos, Johnson, & Chauncey, 2010; Zhou et al., 2010). Computer log files gathered from learners' interaction with online learning environments are the most know and potentially the least obtrusive way of gathering data with regard to learners' SRL behaviour. Such log files are gathered through clickstreams. Clickstreams are also known as click paths, or the route that learners choose when clicking or navigating through an online learning environment. A clickstream is a list of pages visited by a learner, presented in the order the pages are visited (also defined as the 'succession of mouse clicks' that each learners makes). Based on the sequence of the visited pages, researchers attempt to map learners' SRL processes.

1.2 Self-regulated learning and its measurement

As learning in general is seen as an activity performed by learners rather than something happening to them as result of instruction (e.g., Bandura, 1989; Oliver & Trigwell, 2005) it entails a self-regulated process through means of which learners' regulate their behaviour according to the instructional demands (Zimmerman & Schunk, 2001). To be successful learners, learners need to self-regulate their learning. This assumption is evidenced by a substantial body of literature showing scores on self-regulated-learning-related variables to be strongly positive correlated and to have causal relations with scores on performance-related variables (e.g.,

Daniela, 2015; Lin, Coburn, & Eisenberg, 2016). The theoretical conceptualization of SRL evolved from SRL as an aptitude to SRL as an event. The aptitude approach on the one hand sees SRL as in-person, across situations (aggregated over or abstracted from behaviour), and stable from a certain age onwards (e.g., Veenman, 2007; Winne & Perry, 2000). The event approach, on the other hand, conceptualizes SRL as a cyclical process unfolding in roughly three phases (forethought, performance and evaluation) (e.g., Boekaerts, 1992), influenced by variables internal and external to the learner (e.g., Winne & Hadwin, 1998). Additionally, the event approach also sees SRL as covert in nature and so requires inferencing through learners' behaviours and behavioural consequences (e.g., Veenman, Prins, & Verheij, 2003). Although both approaches are still used in research, over the past three decades the event approach gained considerable interest and dominated the investigation of learners' SRL (see: Puustinen & Pulkkinen, 2001).

In line with the shift of conceptualization of SRL, also the conceptualization of measurement approaches to capture it evolved. Measurements shifted from single measurements administered before or after the execution of a task to continuous measurements administered during the execution of the task (Winne & Perry, 2000). The latter type of measurement is referred to as an on-line measurement whereas the former is referred to as the off-line measurement of SRL (Pintrich, 2004). Following the shift in conceptualization of the SRL concept, the use of offline measurements based on learners' perceptions (e.g., self-reports) came under stress (Endedijk et al., 2016). This is mainly because these types of measurements assume learners are capable to predict, reflect, or estimate in general terms (prior or after a task) how they will act in a certain context and subsequently rely on learners' perceptions about their own SRL rather than on the actual account of SRL they exhibit (Veenman, Bavelaar, De Wolf, & Van Haaren, 2014). The interest in sequence data and sequence analysis in SRL taps in into the cyclical nature of SRL and has been particularly fuelled by improvements in technical capabilities. The recording of learning-related behavioural data that are suitable for quantitative analysis has become almost effortless and unobtrusive for the learners in computer-based learning environments (Winne, Nesbit, & Popowich, 2017), making this type of data particularly interesting for both practice and research. Examples of this usefulness are, the mining of theory-based patterns from big data to identifying SRL strategies in massive open online courses (Maldonado-Mahauad, Pérez-Sanagustín, Kizilcec, Morales, & Munoz-Gama, 2018), the finding of traces of SRL in activity streams (Cicchinelli et al., 2018), or the assessment of online learning material and its relation to learners' quantitative behaviour patterns and their effects on motivation and learning performance (Yang, Li, & Xing, 2018). Applications are plentiful.

1.3   Sequence analysis in the field of self-regulated learning

In the field of SRL in general sequence analysis refers to a sequence as an ordering of observable behavioural events preceded and followed by an unknown behavioural state (e.g., Du, Plaisant, Spring, & Shneiderman, 2016; Köck & Paramythis, 2011). Simply put, each change of state is an event, and each event implies a change of state (Müller, Studer, Gabadinho, & Ritschard, 2010). For example, an assumed behavioural state could be reading a content page in an online learning environment, while clicking the calendar tool would be an observable behavioural event that changes the behavioural state of a learner to viewing the calendar page. Through the investigation of ordered observable behavioural events (sequence), researchers try to gain insights in the unknown behavioural states learners are in (Molenaar & Järvelä, 2014). This investigation leads to three types of research questions: (1) questions about the nature of the sequences of the observed events, (2) questions about variables that affect those sequences, and (3) questions about the variables affected by the sequences (Abbott & Tsay, 2000).

To gain a brief insight into the variety of approaches that can be used to handle each of these questions, below we illustrate how the investigation of them can be operationalized. This will be done based on three common phases in the investigation of sequential data in the field of SRL (e.g., Liu et al., 2016; Zhou, 2016). These phases are: (1) the pre-processing phase, (2) the mining and characterization phase, and (3) the analysis phase. Secondly, we provide an illustrative example highlighting (1) the complexity of sequence analysis, (2) the theoretical and methodological choices and considerations to be made, and (3) the reporting of the methods used, illustrating the need for agreed upon frameworks to be able to conduct and report sequence analyses transparently.

1.3.1 Data structure

As the raw computer log file data gathered functions as the input and absolute basis for sequence analysis (Coronel & Morris, 2016), we will start with the description of the data structure of sequence data before elaborating on the different phases of sequence analysis itself. The most common data format of raw computer-log-file data is the time stamped event (TSE) format (Gabadinho, Ritschard, Mueller, & Studer, 2011). A TSE-dataset contains at least three columns: (1) the timestamp of the observable behavioural event, (2) a personal identifier of the learner, and (3) an event name (of the observable behavioural event). Examples of event names are the names of each element in the online learning environment (e.g., discussion form, content page, exercise, etc.), areas on the screen learners clicked, or clicking actions (e.g., Caprotti, 2017; Cicchinelli et al., 2018; Maldonado-Mahauad et al., 2018).

1.3.2 Pre-processing

In a first phase of the sequence-analysis method the raw data is pre-processed (e.g., Zhou, 2016). This pre-processing phase generally consists of two steps. The first step relates to the question: "Is there a need for recoding the raw data?" If there is a need (i.e., deductive approaches) to recode the raw data, this happens using an action library. Such a library specifies the links between the observed events and the recoded, conceptualized concept. Examples of these practices include action libraries based on the strict recoding of events or clusters of events using SRL theories (e.g., Winne et al., 2017). This depends on whether or not the action library is based on think-aloud coding schemes or on pure theoretical conceptualizations (e.g., Bannert et al., 2015; Taub et al., 2017). Another illustration is the coding of computer log files using a tool-related coding scheme (e.g., Lust, 2012; Lust, Vandewaetere, Ceulemans, Elen, & Clarebout, 2011; Siadaty, Gašević, & Hatala, 2016). If there is no need (i.e., inductive approaches) to recode the data, the raw data (as is) can be used (e.g., Kurki, Järvenoja, Järvelä, & Mykkänen, 2017; Van Laer & Elen, 2016). The second step in the pre-processing phase involves assigning an ordered list of events to each learner, resulting in a single sequence per user (Gabadinho et al., 2011). While the chronological ordering of the observed events suffices for the investigation of the sequential nature of such a sequence, the investigation of the temporal characteristics of a sequence will require the calculation and addition of the distance (time) between consecutive events to the sequence. Based on the compilation of a single sequence per learner, sub-sequences and models can be mined and characterized.

1.3.3 Mining and characterization

After the raw data is pre-processed to a single sequence per learner, research questions related to the characteristics of sequences can be investigated. Research questions are plentiful and pertain to the investigation of either whole sequence (β) or sub-sequences (α). A sub-sequence (α) is part of a sequence (β) if the sub-sequence (α) can either directly ($\alpha = <B, D, E, G>$) or indirectly ($\alpha = <C, E, G, E>$) be formed from the sequence ($\beta = <A, C, B, D, E, G, C, E, D, B, G>$) (Zhou et al., 2010). The most common approach to mining and characterizing sequences and sub-sequences is called the algorithmic approach (e.g., Kinnebrew, Loretz, & Biswas, 2013; Perez et al., 2017; Poole, Lambert, Murase, Asencio, & McDonald, 2016). This approach assumes the relation between the different events is unknown and therefore attempts to create meaning from the events that have already occurred by investigating the statistical relationships among them (Breiman, 2001). Efficient algorithms for discovering these characteristics have been proposed in statistical literature. The prominent algorithms are those of Bettini, Wang, and Jajodia (1996), Srikant and Agrawal (1996), Mannila, Toivonen, and Verkamo (1997), Zaki (2001) and Masseglia, Teisseire, and Poncelet (2002). All the algorithms require parameter settings. Examples of these parameter settings are (1) time constraints of the occurrence of an event or sub-sequence, (2) a method for counting the occurrences of events and sub-sequences, and (3) a threshold for the identification of frequently occurring events and sub-sequences. Once the parameters are defined, of-the-shelve software tools makes it possible to apply algorithmic approaches and to identify typical sequences (models), frequent events, and frequent sub-sequences. Common platforms for performing this identification include PROM (process mining workbench), developed by Van der Aalst (2016), SPAM (Sequential Pattern Mining) by Ayres, Flannick, Gehrke, and Yiu (2002), or the TraMineR (trace mining in R) package developed for R-statistics by Gabadinho et al. (2011). An extensive overview of algorithmic tools can be found in Slater, Joksimović, Kovanovic, Baker, and Gasevic (2017). Besides the algorithmic approach described above, there are also other approaches (for an extensive overview see: Poole

et al. (2016)). Examples are theory-driven approach (e.g., Cleary, 2011) which hypothesize the characteristics of sequences and sub-sequences and stochastic approach focussing on whole-sequence modelling (e.g., Biswas, Jeong, Kinnebrew, Sulcer, & Roscoe, 2010; Jeong, Biswas, Johnson, & Howard, 2010). Once sequences and sub-sequences are mined for and characterized they can be used as variables in statistical trials.

1.3.4   Analysis

When investigating sequences in the light of SRL, we may be interested to know how sequences or sub-sequences are impacted by variables internal or external to the learners (e.g., Winne & Baker, 2013). For example when providing an instructional intervention to learners, we might want not only to see the change in learners' learning outcomes but also the change in the occurrence of particular sequences or sub-sequences. Another example might be that we want to compare sequences or sub-sequence of learners with low or high motivation (e.g., Duffy & Azevedo, 2015; Jovanović, Gašević, Dawson, Pardo, & Mirriahi, 2017). In other words we may want to explore which sequences or sub-sequences discriminate most when different groups' sub-sequences or averaged sequence are compared. To answer such questions, various approaches have been proposed for the incorporation of sub-sequences and sequences as dependent variables. The approach of Studer, Mueller, Ritschard, and Gabadinho (2010) consists of measuring the strength of association of each sequence or sub-sequence with the considered covariate and selects the sequence or sub-sequences with the strongest association. The association is measured with the Pearson independence Chi-square. The most discriminant one is the one with the highest Chi-square. Another approach proposed by Kinnebrew et al. (2013) relies on multiple comparisons by t-test statistics between groups based on the considered covariate. The t-test is not used to prove that the groups of sequences differ. Instead, it is employed as a heuristic for identifying more interesting sub-sequences in an exploratory analysis. This is done for example by determining with 95% confidence that a frequent sub-sequence is shown to be different between the groups. Besides these common methods multi-level modelling (e.g., Taub et al., 2016), regression analyses (e.g., Segedy, Kinnebrew, & Biswas, 2015), or Spearman correlation analysis (e.g., Kizilcec, Pérez-Sanagustín, & Maldonado, 2017) are also used.

Besides the investigation of variables influencing sequences, we can also investigate the influence of sequences on another variable. In the field of SRL an example could be the impact of the occurrence of a specific sequence on group performance (e.g., Molenaar & Chiu, 2015). Such research questions investigate the dissimilarities between different sequences (e.g., Abbott & Tsay, 2000; Aisenbrey & Fasang, 2010). These dissimilarities are commonly measured using the optimal matching edit distance. The optimal matching edit distance is defined as the minimal cost of transforming one sequence into the other (e.g., Biemann & Wolf, 2009; Mazon, Rossi, & Toledo, 2014). The transformation operations considered by the optimal matching edit distance are (1) the insertion / deletion cost and (2) a change in the temporal distance resulting in the transformation from one sequence or sub-sequence to another. Event dependent costs can be specified both for the insertion/deletion of an event as well as for a one-unit change in the temporal distance of given events. Both the insertion / deletion and temporal distance cost result in a distance matrix between sequences themselves. This matrix can then be used in classification methods as well as in scaling methods to investigate the relation between various sequences (e.g., Maldonado-Mahauad et al., 2018; Segedy et al., 2015).

1.4   An illustrative example of sequence analysis

Earlier we provided a condensed overview of different choices to be made at each phase of the sequence analysis process. To further illustrate the complexity of sequence analysis, the choices to be made, and the reporting of the methods used we provide an example of a study applying sequence analysis. In the study presented, we investigated the impact of reflection cues on learners' SRL. An event approach to SRL was chosen focussing on SRLs' cyclical, influenceable, and covert nature. SRL was operationalized through learners' learning behaviour and learners' learning outcomes. Two research questions were addressed: the first one investigated the impact of reflection cues on (a) learners' learning behaviour and (b) on learners' learning outcomes. The second research question investigated how learners' learning outcomes related to by learners' learning behaviour. To answer these questions, a 2x2 mixed factorial design was applied and data was gathered

from 60 learners in second chance adult education. Half of the group was exposed to additional cues for reflection; the learners in the control group were not. Learners' behavioural data existed of computer-log-file data gathered through an online learning environment in an ecologically valid setting. Learners' learning outcomes were assessed through cognitive (domain knowledge), motivational (goal orientation), and metacognitive (learning effort and learning confidence) tests and questionnaires. The computer-log-file data gathered had the TSE-format. The event names were actions learners' could perform in the online learning environment (i.e., post in the discussion forum; submit assignment; etc.). As a unit of analysis we used the entire eight week course and instructional stability throughout the eight weeks was described using the instrument of Van Laer and Elen (2018). No validated operationalizations of sequence analysis based on the conceptualizations of the cyclical, influenceable, or covert nature of SRL could be retrieved to direct the operationalization of our investigation. To deal with the issue of the lack of operationalizations, we decided to follow an approach staying as close to the observed data as possible. An inductive rather than a deductive approach was followed to avoid non-transparent alignment between conceptualization and operationalization. In line with this approach, we limited the assumptions made by (1) taking only into account observed overt events, (2) focussing only on the sequential aspects of computer-log-file data, and by (3) not conceptualizing the evolution of SRL through the course of SRL resulting on directly observable patterns via frequent sub-sequences rather than the extraction of behavioural models.

The pre-processing of the data resulted in one (eight week long, +/- 10000 events) sequence of ordered raw behavioural events per learner. No recoding was applied, nor was time between events calculated. In the mining and characterization the TraMineR algorithm (Gabadinho et al., 2011) was used in R-statistics to investigate learners' sequences through the investigation of directly observable patterns via frequent sub-sequences. The identification of frequent sub-sequences was based on (1) the time constraints of the occurrence of events in the observed sub-sequences, (2) a counting method for counting the occurrences of sub-sequences, and on (3) a threshold for the identification of frequently occurring sub-sequences. As only directly observable sub-sequences were targeted, the parameter for the distance between events was set to one, representing that only events directly observed before or after a certain event could be seen as part of a sub-sequence. The counting method chosen was selected arbitrary, based on the occurrence of sub-sequences over the different learners. The frequency threshold was set to 25% meaning that at least 25% of the learners should exhibit the sub-sequence to be counted as frequently occurring. 688 frequent sub sequences were observed.

Next, we investigated frequent sub-sequences' relationship with (1) the condition learners were in (impact of cues on behaviour) and (2) learners' learning outcomes (relations between outcomes and behaviour). In the analysis phase, the frequent sub-sequences were used as dependent variables. For this analysis, chi-square tests were used containing the frequent sub-sequences for discriminating the groups and the variables that defines the groups (condition and learners' learning outcomes). Based on these tests, the effect sizes were calculated using Cramer's V. The Cramer's V expresses the relation between a certain discriminating frequent sub-sequence and the learners' characteristics and is reported in a value between zero and one. The closer to one the higher the relation is. Cohen (1988) refers to small (≤.30), medium (≥.30 and ≤.50), and large (≥.50) effect sizes.

With regard to the first research question dealing with the investigation of the impact of reflection cues on (a) learners' learning behaviour and (b) on learners' learning outcomes, learners in the experimental condition were shown to make significantly more use of sub-sequences consisting of events related to assignments and tasks, communication, and assessment. Furthermore, both conditions showed a significant increase in domain knowledge and learning confidence and a decrease in performance goal approach. Learners in the experimental condition who received cues for reflection scored significantly higher on performance goal approach compared to the learners in the control condition. As for the interaction effect between time and condition, learners in the experimental condition scored significantly higher for performance avoidance approach compared to their counterparts. This result was unexpected in the light of the aim of the study (Van Laer et al., 2018). Finally, with regard to the second research question dealing with the investigation of how learners' learning outcomes related to by learners' learning behaviour, it became clear that changes in learning behaviour seemed to be linked to learning outcomes (performance avoidance approach). Results showed that differences in learners' learning behaviour were observed when learners had different performance avoidance approach scores.

1.5 Towards tangible proof of progress

As research aims at either at building or testing theory, the research cycle moves from description, to explanation, to testing with repeated iterations through this cycle (Van der Merwe, 2013). Throughout this iterative process, descriptive models are expanded into explanatory frameworks that are tested against reality until they are eventually developed into theories as research study builds upon research study. The result is to validate and add confidence to previous findings, or else invalidate them and force researchers to develop more valid or more complete theories (Meredith, 1993). In this way both (1) theoretical conceptualizations of the theory under investigation and their (2) operationalization through measurements are continuously updated and refined. As illustrated throughout the different paragraphs presented above, different operationalizations of sequence analysis can be made. The illustrative example has shown one of these operationalizations.

To be able to monitor methodologies' evolution towards tangible proof of progress and so secure the iterative research cycle, the literature on advances in research methodology (e.g., Beach & Pedersen, 2013; Lupia & Alter, 2014) proposes three indications of such an evolution. The first one is the transparency of the method applied (Moravcsik, 2014). Transparently reported methods permit scholars to assess research and to communicate with one another. Unless other scholars can examine evidence, parse the analysis, and understand the processes by which evidence and theories were chosen, why should they trust and thus expend the time and effort to scrutinize, critique, debate, or extend existing research? As demonstrated earlier, a lot of explorative work on the use of sequence analysis has been done, yet most of the studies do not seem to report in detail on the different phases of sequence analysis or on the parameter settings involved in each of them, hampering a thorough study of the method applied. When literature on the investigation of SRL through sequence analysis reports on the log file data structure (e.g., Biswas, Roscoe, Jeong, & Sulcer, 2009; Duffy & Azevedo, 2015; Lazakidou & Retalis, 2010), it often does this through elaborating on the events traced: clicks, pages, specific cognitive or metacognitive activities, and so on. Despite the information on the events traced, additional information on the structure of the data, such as the timestamp interval or type of timestamps, session identifiers, etc., is hardly provided. This information is important to distinguish which pre-processing steps are possible or desirable (e.g., calculation of time between events, grouping of learners or individuals, etc.). With regard to this pre-processing phase, in the best cases researchers acknowledge they developed a set of filters or recoding algorithms to remove irrelevant information from the raw log files, with the aim of presenting the relevant information in a compact format that is suitable for further analysis (e.g., Jeske, Backhaus, & Stamov Roßnagel, 2014; Paans, Molenaar, Segers, & Verhoeven, 2018). Nonetheless they hardly ever elaborate on which information was discarded and what made the researcher assume this information could be classified as irrelevant. When for example action libraries are used (e.g., Bannert et al., 2014; Goldberg et al., 2014) researchers elaborate on the coding scheme, but lack to state how the 'raw' events are recoded and what the reliability of this recoding was like. Without this information it is impossible to distinguish which coding scheme is most reliable and works best for what data. In line with this, no studies seem to be available which argue for the selection of a certain coding scheme or elaborate on why a certain coding scheme is preferable over another. With regard to the mining and characterization of sequences, most of the current literature seems to indicate which algorithms are used to identify or mine (frequent) sub-sequences. Nonetheless, the authors rarely seem to address the assumptions underlying the algorithm used (e.g., Balderas, Dodero, Palomo-Duarte, & Ruiz-Rube, 2015; Lan & Lu, 2017) or the procedure followed to select the appropriate algorithm (e.g., Kizilcec et al., 2017; Maldonado-Mahauad et al., 2018), let alone the parameter settings applied when mining for sub-sequences. The same is the case when the (frequent) sub-sequences are used as dependent or independent variables. The analysis methods used to answer similar research questions vary from researcher to researcher (Ahmadpour & Khaasteh, 2017; Cerezo, Esteban, Sánchez-Santillán, & Núñez, 2017; Chen, Breslow, & DeBoer, 2018). Traditional cluster analysis and predictive apriori algorithms are used to identify sets of successful learner and environmental characteristics impacting performance, without any explanation of why a certain approach might be considered superior to another. The multitude of methods and the observation of the lack of transparency make the studies unreplicable.

The second characteristic relates to the availability of comparative research designs (e.g., Bureau & Salomonsen, 2012; Peterson, 2005). Comparison is one of the most powerful tools used in intellectual inquiry,

since an observation made repeatedly is given more credence than a single observation. Put simply, as argued by Mills, Van de Bunt, and De Bruijn (2006) the main goal of comparative research is to search for or identify variance or similarity. Although there is quite some comparative research on the measurement of SRL, the majority of it focusses at best on the comparison of online behaviour event measurements (i.e., sequence analysis) with offline perception event measurements (i.e., self-reports) (e.g., Cho & Yoo, 2017; Hadwin, Nesbit, Jamieson-Noel, Code, & Winne, 2007). Even at the most basic level of comparison, namely the use of different coding schemes to recode the 'raw' event captured in log files, there seems to be hardly any evidence on which coding scheme results in the most accurate results under which conditions (Azevedo, 2014). Although there are useful summaries of approaches and tools (e.g., Slater et al., 2017) as well as ample ideas on how to apply sequence analysis (e.g., Azevedo et al., 2010; Winne, 2018; Winne et al., 2017), no discussion of different sequence analysis methods could be found in the field of SRL. Based on this observation, research comparing different sequence analysis approaches for SRL seems to be missing. This would lead to the identification of commonalities and differences between methods adding to the validation of the method.

The third and final characteristic is the application of a method in empirical ecologically valid settings (e.g., Chambless & Ollendick, 2001; Rotter, 1954). Only when methods can be applied in different contexts and situations, they can propel and more importantly validate investigations. Although there have been attempts in the field of self-regulated learning to operationalize insights drawn from experimental settings in ecologically valid empirical contexts, these attempts are mainly based on a mixture of insights obtained from the experimental setting, accompanied by a data-driven approach to overcome the gaps left by the experimental approach (e.g., parameter settings, coding events, identification of sub-sequences) (e.g., Hsu, 2018; Ifenthaler, Gibson, & Dobozy, 2018; Taub, Azevedo, Bradbury, Millar, & Lester, 2018). No clear attempts to apply and transfer insights between settings seem to be made so far.

In summary, it becomes clear that a lot of work still needs to be done. It seems that none of the three indications for tangible proof of progress already has been achieved for the use of sequence analysis in the field of SRL. In line with this finding already in 2014, Roger Azevedo (2014) pointed out that researchers investigating sequence data recoded the data different, made diverse statistical and theoretical assumptions regarding the data collected, and that too easily inferences were drawn from the sequential and temporal unfolding data. His call for action was in vane and repeated multiple times (e.g., Molenaar & Järvelä, 2014; Winne et al., 2017), supported by the expression of the need for standards and frameworks to align the investigation of sequence data in the field of SRL (e.g., Bannert et al., 2015). Partly, the aim of this manuscript is to add to the body of literature calling for standards, protocols, and frameworks that can be tested and validated. By providing general guidelines contributing to a framework for the use and reporting of sequence analysis for SRL this manuscript aims to propel the establishment of sequence analysis as a research method.

1.6   Problem statement

As illustrated in the current section, sequence analysis in the field of SRL is an umbrella term covering a large variety of approaches. With regard to the log file data format and the pre-procession phase it often seems unclear how researchers devise and deploy the data scrubbing, cleansing, recoding, or the cleaning processes (e.g., Clarke, 2016; Müller, Naumann, & Freytag, 2003; Rahm & Do, 2000). With regard to the mining and characterization phase and to the analysis phase hardly any explicit references seem to be made to the basis of specific decisions. Current research makes it hard to distinguish which parameter settings were derived from literature or which ones were set arbitrary, nor why this is the case. No arguments are given on why a certain approach is considered above another one (e.g., Poole et al., 2016; Stark & Vedres, 2012). The multitude of approaches and considerations does not yet seem to be condensed into a transparent methodological framework for sequence analysis, nor do these approaches seem to contribute yet to tangible proof of progress in the investigation of learners' SRL using sequence analysis based on computer log files. Without such a framework it is impossible to test, falsify or modify approaches to sequence analysis and so spark the investigation of learning through sequence analysis. In the next section we propose a set of guidelines functioning as a potential starting point for the construction of a transparent methodological framework to communicate approaches of sequence analysis in the field of SRL.

2. Guidelines on the use of sequence analysis in the field of self-regulated learning

As illustrated throughout the introductory section of this manuscript, the operationalization of sequence analysis to investigate learners' SRL using computer log files is shaped through many practical choices and theoretical assumptions. Whereas in the previous sections we highlighted the need for a methodological framework, in what follows we propose a set of guidelines functioning as a potential starting point for the construction of such a framework. We offer guidelines in two main areas. The first one relates to the alignment of the conceptualization and the operationalization of the different components of SRL. The second one relates to the enactment of the operationalization of the selected sequence analysis approach.

2.1. Alignment of conceptualization and operationalization

Singleton, Straits, and Straits (1999) see alignment of conceptualization and operationalization as one of the key features to scientific and methodological success. They refer to the process of conceptualization as the act of defining the different components of a phenomenon under investigation and to operationalization as the practical result of the conceptualization act. In line with this definition, we explore the operational impact of the current conceptualization of SRL. As presented in the introduction, current conceptualizations of SRL focus on its cyclical, influenceable, and covert nature (e.g., Winne & Hadwin, 1998). In what follows we relate these three general conceptions to practical consequences in the operationalization of the selected sequence analysis approach chosen. Even when these three conceptions are not at the basis of the conceptualization of SRL, the conceptions below might shed light on the relation between on the one hand the conceptualization of the phenomenon under investigation and the selected operationalization of sequence analysis.

2.1.1. The cyclical nature of self-regulated learning

The idea of SRL phases unfolding in different cyclical phases raises questions concerning (1) the dynamics of these cyclical SRL process, (2) the sequential patterns within it, as well as (3) the development of the cycle over time. Each of these questions bring the notion of sequentiality and temporality to the discourse on SRL (e.g., Molenaar & Järvelä, 2014). While in the literature on SRL the 'temporal' and 'sequential' notion is often used intertwined (Knight, Wise, & Chen, 2017), literature on sequence analysis from other fields of research makes a clear distinction. Temporality refers to the passage of elapsed time and comes with a collection of related concepts such as duration, rate, and acceleration (Blikstein, 2011; Haythornthwaite & Gruzd, 2012). Sequentiality is used to refer to the order of events and transitions between different events, without explicit reference needed to duration or passages of time (Biswas et al., 2010; Halatchliyski, Hecking, Goehnert, & Hoppe, 2014). As a first consideration, when for example incorporating only sequential characteristics of SRL, the construction of a single sequence per learner in the pre-processing phase will consist of the chronological ordering of events, assuming time between each events' timestamp of secondary interest. When in contrast also temporal characteristics of SRL are taken into account, the construction of learners' single sequences will include the calculation of the time between the consecutive events' timestamps and the inclusion of these calculations in further analyses. The latter poses additional conceptual questions to the status of this calculation. Does it for example represents a single hidden unknown state or is it instead an indication of involvement with the environment? A second consideration relates to the developmental characteristic of the behaviour observed. When SRL-development over time is assumed (e;g., Andrade & Evans, 2015; Huang, Klein, & Beck, 2017) a whole-sequence approach might be preferred over a sub-sequence approach that does not consider such a development (in the time frame of investigation) (Winne & Hadwin, 1998). Both will affect further analyses. As demonstrated above while briefly reviewing the conceptualization of the cyclical nature of SRL, a clear link between (1) the conceptualization of the sequential and temporal characteristics of SRL and its practical operationalization and (2) the developmental characteristics of SRL over time and its operationalization in practice seem necessary to be able to study the sequence analyses applied.

2.1.2. The influenceable nature of self-regulated learning
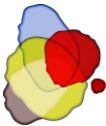
With regard to how SRL comes to be, recent event theories regard SRL as influenced by variables internal and external to the learner (Veenman, Van Hout-Wolters, & Afflerbach, 2006). In general research identifies three major sets of internal variables influencing SRL: cognitive (e.g., Zimmerman, 1986, 1990, 1998; Zimmerman & Pons, 1986), metacognitive (Borkowski, Carr, Rellinger, & Pressley, 1990; Pressley, Levin, & McDaniel, 1987) and motivational variables (e.g., Butler & Winne, 1995; Schraw, Crippen, & Hartley, 2006; Schraw & Moshman, 1995; Zimmerman, 2000). A substantial body of literature identifies external variables at different grainsize levels influencing learners' SRL. Dignath and Büttner (2008) for example point in their meta-analysis that (1) instruction of cognitive strategies (i.e., rehearsal, elaboration, and organizational strategies) affected learners' SRL significantly. The same was observed for (2) instruction of metacognitive strategies (i.e., planning, monitoring, and evaluation), (3) promoting metacognitive reflection, and (4) instruction of motivation strategies. Another example is the literature review of Van Laer and Elen (2016) identifying seven attributes of learning environments that support learners' SRL. The combinations of the abovementioned internal and external variables make up the timeframe in which SRL needs to be investigated. Under this conceptualization, each change in either variables internal and / or external to the learner will influence learners' SRL (e.g., Greene & Azevedo, 2007; Winne & Hadwin, 1998). Thus without the appropriate timeframe in which to investigate learners' SRL insights might be hard to gather. So, the main consideration with regard to the influenceable nature of SRL relates to the unit of analysis. The size of the allowed timeframe affects the operationalization of sequence analysis for example through the parameter settings while mining and characterizing sequence and sub-sequence. When for example both internal and external variables are regarded as stable throughout the investigative trial the timeframe might stretch over the entire trial. If in contrast the variables are assumed to be variable at a certain rate, the timeframe might want to match this rate as much as possible. As illustrated before the characterization of sequences and sub-sequences relies amongst others on specifications with regard to the time constraints of the occurrence of an event or sub-sequence. The time dimension raises questions concerning the maximal distance between events (Molenaar & Järvelä, 2014). Moreover, we may consider two or more events form a relevant sequence or sub-sequence only if they occur within a given distance of each other (maximal timespan). For example when they occur in a time window in which both variables within and external to the learner are assumed to be constant. For example completing an exercise right after viewing a content related page might not mean the same as completing that same exercise twenty events after viewing the content related page (e.g., Du et al., 2016; Jovanovic, Pardo, Mirriahi, Dawson, & Gašević, 2017). To conclude, it seems that to be able to study the sequence analyses applied, the operationalization of the unit of analysis (e.g., through parameter settings) as conceptualized through the influenceable nature of SRL needs to be elaborated. Elaborating on the relation between unit of analysis and parameter settings allows us to assess the suitability of the decisions made.

2.1.3. *Covert nature of self-regulated learning*

Current conceptualizations assume SRL operates at different levels of the cognitive system and so regulates lower order cognitive processes that, in turn, shape learners' overt cognitive behaviour (Roth, Ogrin, & Schmitz, 2016). This conceptualization results in the assumption that SRL occurring in each of the SRL phases cannot be directly observed as it manifests in overt cognitive behaviours (Williamson, 2015) and through behavioural consequences like learners' learning outcomes (Veenman & Alexander, 2011). For instance, when a learner recalculates the outcome of a mathematical equation, it is assumed that a SRL monitoring or evaluation process must have preceded this overt cognitive activity of recalculation. As illustrated in the introductory section, conceptualization of the covert nature of SRL can be based on the relation between the overt behavioural events with SRL-influencing constructs (e.g., tool-use, engagement, etc.) or directly with SRL-related activities (e.g., goal-setting and planning, monitoring, etc.) (e.g., Azevedo et al., 2015; Bannert et al., 2015; Lust, 2012). Depending on the conceptualization of (a) the covert nature of SRL and (b) how this nature can be uncovered through the overt behavioural events observed through computer log files, a link will be constituted with the operationalization of this covert nature through the establishment of action libraries (Zhou, 2016). When an action library is used, overt behavioural events (or sets of events) are recoded into more meaningful learner behaviour or SRL behaviour. Although this practice is common when following a deductive research approach, action libraries often have a different level of granularity as they are developed through other online event measurements (i.e., think aloud trials) (Azevedo, 2014). Comparing the micro-level approach followed through using computer log files with the codes abstracted from think aloud

trails might be problematic given the different grain-size (e.g., Al Mamun, Lawrie, & Wright, 2017). In summary, when operationalizing the covert nature of SRL via recoding data using different grainsizes (i.e., computer log files and think aloud data), clearly the relationship between observed behaviour and assigned codes needs to be made explicit and communicated transparently.

2.2.   Enactment of the operationalization of sequence analysis

Proctor, Powell, and McMillen (2013) see enactment as the step following the conceptualization and operationalization of a research method. They identify enactment as the systematic application of the operationalized concepts. As illustrated before, the quest for tangible proof of progress lies in transparently reported methods and procedures permitting scholars to assess research and to communicate with one another (e.g., Beach & Pedersen, 2013; Lupia & Alter, 2014). With regard to the enactment of the operationalization of the selected sequence approach, we focus on two components: systematic account of the operationalization and transparent parameter settings.

2.2.1.   Systematic account of the operationalization

The operationalization of sequence analysis starts from gathered data with a specific structure and unfolds roughly in three phases. In the first phase, the pre-processing phase, a single sequence per learner is constructed. In the second phase, sequences and frequent sub-sequences are mined and characterized. Identified sequences and sub-sequences can function as either dependent or independent variables. Although general approaches such as the one described above have been usefully proposed by for example Zhou (2016) and Liu et al. (2016), current research fails to go much beyond the vagueness level of this general approach. As demonstrated a multitude of decisions need to be made in the chain of sequence analysis (Roll & Winne, 2015). Keeping systematicity and transparency in mind, detailed accounts of each of these phases would increase replicability (e.g., Moravcsik, 2014). Systematic accounts of the enactment of the operationalization of sequence analysis might start with a description of the raw data gathered. This not only means reporting on the environment in which the data was gathered, but also the actual structure of the dataset extracted, including for example the database structure. In the pre-processing phase systematicity and transparency might be accomplished through elaborating (among others) on the data cleaning process, the recoding procedure (if applicable), and the transformations applied (if applicable). With regard to the mining and characterization of the sequences and sub-sequences this might be done through a clear account of the different steps taken in the mining and characterization process, including for example a detailed explanation of the algorithm used and the parameters set. Finally in the analysis phase transparency and systematicity might be accomplished by the presentation of the output format of the previous phase and the analysis approach chosen (with its key figures).

2.2.2.   *Transparent parameter settings*

It is clear that the conceptualization of a theory cannot account for each variation in operationalization nor for the justification of each parameter setting (Bannert et al., 2017). Nonetheless, decisions need to be made to derive useful approaches. Regardless of the inductive or deductive approach to the conceptualization of SRL, transparency and courage on the part of the researchers is essential to report in detail which parameter settings are derived from theory and which are arbitrary (e.g., Tsai, Shen, & Tsai, 2011). The degree to which this is possible on either side is irrelevant to the argument, as long it is clear which parameter settings are used for what reason or considering what assumption. An example of such a practice was presented in the illustrative case. No theoretical evidence could be found to determine the frequency threshold for identifying frequent sub-sequences. In this case, it was reported that an arbitrary cut-off was set at 25%.

3.   Implications and conclusions

Although the use of sequence analysis in the field of self-regulation is one of the last decade, a lot of valuable work has been done to propel the investigation of learners' SRL through sequence analysis. Despite

these efforts, no methodological framework seems to be available for the systematic application and reporting of sequence analysis in the field of SRL. Because of the lack of such a framework, tangible proof of progress is difficult to achieve and so the evolution of sequence analysis to investigate SRL seems to be hampered. To illustrate the need for such a methodological framework we provided in the introduction of this manuscript a brief overview of the variability of current operationalizations and illustrated one such approach. From the pre-processing phase onwards, over the mining and characterization phase, up to the use of the identified sequences and sub-sequences as dependent and independent variables, a multitude of conceptual and operational choices need to be made. Therefore, in this manuscript we aimed to foster discussion on a methodological framework for the application of sequence analysis in the field of SRL which would make replication, falsification, and validation possible. To do so, in addition to the case built in the introduction section, we proposed in the previous section a set of guidelines functioning as a potential starting point for the construction of a framework. These guidelines were centred on two key areas. The first area focussed on the alignment of the conceptualization of the different components of SRL and the operationalization of the selected sequence analysis approach (e.g., Singleton et al., 1999). The other area focussed on the enactment of the operationalization of the selected sequence analysis approach (e.g., Proctor et al., 2013). With regard to the former four guidelines were proposed relating to the current conceptualization of SRL that relate to: (1) the sequential and temporal characteristics of SRL; (2 the development through time of SRL; (3) the unit of analysis imposed by the factors influencing SRL; (4) the matching-granularity as linked to the covert nature of SRL. With regard to the enactment of the operationalization of the sequence analysis approach two guidelines are proposed related to: (1) the systematic account of the operationalization; and (2) the transparent communication of parameter settings. Although this manuscript does not pretend to provide solutions nor to be exhaustive with regard to possible approaches to sequences analysis in the field for SRL, on the one hand it highlights the need for a transparent and systematic methodological approach. On the other hand, it also identifies guidelines that might function as a basis for further construction of such a methodological framework for the use of sequence analysis in the field of SRL. Keeping in mind the nature of the guidelines provided, we might wonder whether the guidelines are simply 'common sense' and applicable to many other research methods. The latter is certainly the case, yet guidelines have not been constructed before for the investigation of sequence analysis for SRL. With regard to the 'ordinariness' of the guidelines provided in this manuscript, there is an abundance of systematic methodological literature reviews (e.g., Kallio, Pietilä, Johnson, & Kangasniemi, 2016; Kelly, Lesh, & Baek, 2014; Mertens, 2014) illustrating that guidelines similar to the ones suggested in this manuscript might be beneficial for a broad range of research methods and also that without such guidelines, the quest for tangible proof of progress is more than likely to be unsuccessful.

3.1. Implications

When a methodological framework is in place the investigation of time-related characteristics in SRL, using sequence analysis might evolve to (1) transparent methods, (2) comparative studies, and (3) empirical and ecological applications, supporting both research and practice. Such a framework enables researchers to use the framework to describe and compare current approaches to sequence analysis. A solid description of these approaches, their reproduction and validation first in similar, later in different contexts. The former will allow us to apply the insights gathered not only under very strict conditions in one particular situation but is likely to foster empirical and ecologically valid trails. The latter would be useful for researchers and practitioners using sequence analysis for example to inform the design of their course. By having a methodological framework at their disposal, selecting the most appropriate sequence analysis approach for their needs is facilitated. The sooner we are able to compare, validate, and establish sequence analysis methods, the more quickly we can make progress in the investigation of SRL through learners' learning behaviour.

3.2. Further directions

As it was not our aim to present a fully developed methodological framework for the use of sequence analysis in the field of SRL, in future investigations it might be interesting to further detail the conceptual assumptions related to the investigation of time-related characteristics of SRL and their relation to

methodological operationalizations. This could be done by incorporating more theoretical research on the investigation of self-regulated learning and extract the possible methodological consequences for the operationalization of the theoretical conceptualizations proposed. Another avenue might be the integration of non-content-related research on sequence analysis to further investigate the operational possibilities of the method to further investigate the conceptual assumptions made by choosing a particular approach over another. By doing so a protocol, standard, or framework can be established as a method for the use of sequence analysis which can then be tested, validated and modified to further optimize the use of sequence analysis for the investigation of SRL.

3.3. Conclusions

First, the manuscript built a case for the need of a methodological framework for the application of sequence analysis in the field of self-regulated learning. Secondly, it provided a ground for further discussion on the construction of such methodological framework, raising questions about both the conceptualization and the operationalization of sequence analysis in the field of SRL. Additionally it provides guidelines and possible directions supporting sequence analysis as a method in the field of SRL. Applying sequence analysis as a method in the field of SRL in a more systematic and transparent way might support the development of the method towards more transparency, comparative studies, and empirical and ecological applications, supporting both research and practice. As demonstrated throughout the manuscript it seems that it is not the amount of data gathered that will help us gain insights, but rather the way we analyse them and the thoroughness of that analysis. This manuscript by no means implies that the overview of operationalizations is exhaustive or complete, nor does is pretend to provide a best practice or example through the illustrative example. Instead, it aimed to provide a transparent and verifiable framework to discuss the method we see as potentially powerful for investigating learners' SRL. Such a framework challenges the assumptions made, approaches taken, and thus propels the investigation of learners' SRL through computer log files to new heights.

In conclusion, the guidelines proposed and their underlying call for transparency and systematicity through the construction of a methodological framework for the use of sequence analysis in the field of SRL can potentially transcend the use of sequence analysis for computer log files and go as far as to other log file methods investigating the temporal and sequential nature of phenomena. As the literature on the investigation of for example eye movement (e.g., Kiefer, Giannopoulos, Raubal, & Duchowski, 2017; Lorigo et al., 2008) or skin conduction (e.g., El-Sheikh, 2007; Haufler et al., 2017) log files seems to experience similar issues, also these fields of research might benefit from the general guidelines formulated in the manuscript presented (e.g., Kuhn, 2012). Although the focus in this manuscript is only on one rather specific approach to sequence analysis, the quest for transparency and systematicity is one that relates to all investigations, especially when it comes to new complex methods that require abundant data processing in order to make them meaningful.

Key points

- Substantial increase in the use sequence-analysis methods for investigating self-regulated learning.
- There is hardly any tangible proof of progress of sequences analysis as a method.
- Alignment of conceptualization and operationalization and transparent enactment is desirable.
- When such a framework is in place, sequence analysis is more likely to grow as a method.

Acknowledgments

References

Abbott, A., & Tsay, A. (2000). Sequence analysis and optimal matching methods in sociology: Review and prospect. *Sociological methods & research, 29*(1), 3-33. doi: 10.1177/0049124100029001001

Ahmadpour, Z., & Khaasteh, R. (2017). Writing Behaviors and Critical Thinking Styles: The Case of Blended Learning. *Khazar Journal of Humanities & Social Sciences, 20*(1).

Aisenbrey, S., & Fasang, A. E. (2010). New life for old ideas: The" second wave" of sequence analysis bringing the" course" back into the life course. *Sociological methods & research, 38*(3), 420-462.

Al Mamun, M. A., Lawrie, G., & Wright, T. (2017). *Factors affecting student engagement in self-directed online learning module.* Paper presented at the Proceedings of The Australian Conference on Science and Mathematics Education

Andrade, M. S., & Evans, N. W. (2015). Developing Self-Regulated Learners. *ESL readers and writers in higher education: Understanding challenges, providing support*, 113.

Antunes, C. M., & Oliveira, A. L. (2001). *Temporal data mining: An overview.* Paper presented at the KDD workshop on temporal data mining.

Ayres, J., Flannick, J., Gehrke, J., & Yiu, T. (2002). *Sequential pattern mining using a bitmap representation.* Paper presented at the Proceedings of the eighth ACM SIGKDD international conference on Knowledge discovery and data mining.

Azevedo, R. (2014). Issues in dealing with sequential and temporal characteristics of self-and socially-regulated learning. *Metacognition and Learning, 9*(2), 217-228.

Azevedo, R., & Hadwin, A. F. (2005). Scaffolding self-regulated learning and metacognition–Implications for the design of computer-based scaffolds: Springer.

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. doi: 10.1080/00461520.2010.515934

Azevedo, R., Taub, M., & Mudrick, N. (2015). Technologies supporting self-regulated learning. *The SAGE encyclopedia of educational technology*, 731-734.

Balderas, A., Dodero, J. M., Palomo-Duarte, M., & Ruiz-Rube, I. (2015). A domain specific language for online learning competence assessments. *International Journal of Engineering Education, 31*(3), 851-862.

Bandura, A. (1989). Human agency in social cognitive theory. *American psychologist, 44*(9), 1175.

Bannert, M., Molenaar, I., Azevedo, R., Järvelä, S., & Gašević, D. (2017). *Relevance of learning analytics to measure and support students' learning in adaptive educational technologies.* Paper presented at the Proceedings of the Seventh International Learning Analytics & Knowledge Conference.

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.

Bannert, M., Sonnenberg, C., Mengelkamp, C., & Pieger, E. (2015). Short-and long-term effects of students' self-directed metacognitive prompts on navigation behavior and learning performance. *Computers in Human Behavior, 52*, 293-306.

Beach, D., & Pedersen, R. B. (2013). *Process-tracing methods: Foundations and guidelines*: University of Michigan Press.

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.* Paper presented at the Proceedings of the Fifth International Conference on Learning Analytics And Knowledge.

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. doi: 10.1007/s11409-015-9134-6

Bettini, C., Wang, X. S., & Jajodia, S. (1996). *Testing complex temporal relationships involving multiple granularities and its application to data mining.* Paper presented at the Proceedings of the fifteenth ACM SIGACT-SIGMOD-SIGART symposium on Principles of database systems.

Biemann, T., & Wolf, J. (2009). Career patterns of top management team members in five countries: An optimal matching analysis. *The International Journal of Human Resource Management, 20*(5), 975-991. doi: 10.1080/09585190902850190

Biswas, G., Jeong, H., Kinnebrew, J. S., Sulcer, B., & Roscoe, R. (2010). Measuring self-regulated learning skills through social interactions in a teachable agent environment. *Research and Practice in Technology Enhanced Learning, 5*(02), 123-152.

Biswas, G., Roscoe, R., Jeong, H., & Sulcer, B. (2009). *Promoting self-regulated learning skills in agent-based learning environments.* Paper presented at the Proceedings of the 17th international conference on computers in education.

Blikstein, P. (2011). *Using learning analytics to assess students' behavior in open-ended programming tasks.* Paper presented at the Proceedings of the 1st international conference on learning analytics and knowledge.

Boekaerts, M. (1992). The adaptable learning process: Initiating and maintaining behavioural change. *Applied Psychology, 41*(4), 377–397. doi: 10.1111/j.1464-0597.1992.tb00713.x

Bonin, F., Vogel, C., & Campbell, N. (2014). *Social sequence analysis: temporal sequences in interactional conversations.* Paper presented at the Cognitive Infocommunications (CogInfoCom), 2014 5th IEEE Conference on.

Borkowski, J. G., Carr, M., Rellinger, E., & Pressley, M. (1990). Self-regulated cognition: Interdependence of metacognition, attributions, and self-esteem. *Dimensions of thinking and cognitive instruction, 1*, 53-92.

Bourbonnais, S., Hamel, E. B., Lindsay, B. G., Liu, C., Stankiewitz, J., & Truong, T. C. (2006). Method, system, and program for merging log entries from multiple recovery log files: Google Patents.

Breiman, L. (2001). Statistical modeling: The two cultures (with comments and a rejoinder by the author). *Statistical science, 16*(3), 199-231.

Bureau, V., & Salomonsen, H. H. (2012). Comparing Comparative Research Designs.

Butler, D. L., & Winne, P. H. (1995). Feedback and Self-Regulated Learning: A Theoretical Synthesis. *Review of Educational Research, 65*(3), 245-281. doi: 10.2307/1170684

Caprotti, O. (2017). Shapes of educational data in an online calculus course. *Journal of Learning Analytics, 4*(2), 76-90.

Cerezo, R., Esteban, M., Sánchez-Santillán, M., & Núñez, J. C. (2017). Procrastinating Behavior in Computer-Based Learning Environments to Predict Performance: A Case Study in Moodle. *Frontiers in Psychology, 8*, 1403.

Chambless, D. L., & Ollendick, T. H. (2001). Empirically supported psychological interventions: Controversies and evidence. *Annual Review of Psychology, 52*(1), 685-716. doi: 10.1146/annurev.psych.52.1.685

Chen, X., Breslow, L., & DeBoer, J. (2018). Analyzing productive learning behaviors for students using immediate corrective feedback in a blended learning environment. *Computers & Education, 117*, 59-74.

Cho, M.-H., & Yoo, J. S. (2017). Exploring online students' self-regulated learning with self-reported surveys and log files: a data mining approach. *Interactive Learning Environments, 25*(8), 970-982.

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.

Clarke, R. (2016). Big data, big risks. *Information Systems Journal, 26*(1), 77-90. doi: 10.1111/isj.12088

Cleary, T. J. (2011). Emergence of self-regulated learning microanalysis. *Handbook of self-regulation of learning and performance*, 329-345.

Coronel, C., & Morris, S. (2016). *Database systems: design, implementation, & management*: Cengage Learning.

Daniela, P. (2015). The relationship between self-regulation, motivation and performance at secondary school students. *Procedia-Social and Behavioral Sciences, 191*, 2549-2553. doi: 10.1016/j.sbspro.2015.04.410

Dignath, C., & Büttner, G. (2008). Components of fostering self-regulated learning among students. A meta-analysis on intervention studies at primary and secondary school level. *Metacognition and Learning, 3*(3), 231-264. doi: 10.1007/s11409-008-9029-x

Du, F., Plaisant, C., Spring, N., & Shneiderman, B. (2016). *EventAction: Visual analytics for temporal event sequence recommendation.* Paper presented at the Visual Analytics Science and Technology (VAST), 2016 IEEE Conference on.

Duffy, M. C., & Azevedo, R. (2015). Motivation matters: Interactions between achievement goals and agent scaffolding for self-regulated learning within an intelligent tutoring system. *Computers in Human Behavior, 52*, 338-348.

El-Sheikh, M. (2007). Children's skin conductance level and reactivity: Are these measures stable over time and across tasks? *Developmental Psychobiology, 49*(2), 180-186.

Endedijk, M. D., Brekelmans, M., Sleegers, P., & Vermunt, J. D. (2016). Measuring students' self-regulated learning in professional education: bridging the gap between event and aptitude measurements. *Quality & Quantity, 50*(5), 2141-2164.

Gabadinho, A., Ritschard, G., Mueller, N. S., & Studer, M. (2011). Analyzing and visualizing state sequences in R with TraMineR. *Journal of Statistical Software, 40*(4), 1-37.

Goldberg, B., Sottilare, R., Roll, I., Lajoie, S., Poitras, E., Biswas, G., . . . Long, Y. (2014). Enhancing self-regulated learning through metacognitively-aware intelligent tutoring systems: Boulder, CO: International Society of the Learning Sciences.

Greene, J. A., & Azevedo, R. (2007). A Theoretical Review of Winne and Hadwin's Model of Self-Regulated Learning: New Perspectives and Directions. *Review of Educational Research, 77*(3), 334-372. doi: 10.3102/003465430303953

Hadwin, A. F., Nesbit, J. C., Jamieson-Noel, D., Code, J., & Winne, P. H. (2007). Examining trace data to explore self-regulated learning. *Metacognition and Learning, 2*(2-3), 107-124.

Halatchliyski, I., Hecking, T., Goehnert, T., & Hoppe, H. U. (2014). Analyzing the main paths of knowledge evolution and contributor roles in an open learning community. *Journal of Learning Analytics, 1*(2), 72-93.

Haufler, A. J., Lewis, G. F., Davila, M. I., Westhelle, F., Gavrilis, J., Bryce, C. I., . . . McDaniel, W. (2017). Biobehavioral Insights into Adaptive Behavior in Complex and Dynamic Operational Settings: Lessons learned from the Soldier Performance and Effective, Adaptable Response (SPEAR) Task. *Frontiers in medicine, 4*, 217.

Haythornthwaite, C., & Gruzd, A. (2012). *Exploring patterns and configurations in networked learning texts.* Paper presented at the System Science (HICSS), 2012 45th Hawaii International Conference on.

Hine, C. (2011). Internet research and unobtrusive methods. *Social Research Update*(61), 1.

Hsu, T.-C. (2018). Behavioural sequential analysis of using an instant response application to enhance peer interactions in a flipped classroom. *Interactive Learning Environments, 26*(1), 91-105.

Huang, N., Klein, M., & Beck, A. (2017). Measuring student teachers development of metacognition and self-regulated learning in professional dialogue. *ECER 2017*.

Ifenthaler, D., Gibson, D., & Dobozy, E. (2018). Informing learning design through analytics: Applying network graph analysis. *Australasian Journal of Educational Technology, 34*(2).

Jeong, H., Biswas, G., Johnson, J., & Howard, L. (2010). *Analysis of productive learning behaviors in a structured inquiry cycle using hidden markov models.* Paper presented at the Educational Data Mining 2010.

Jeske, D., Backhaus, J., & Stamov Roßnagel, C. (2014). Self-regulation during e-learning: using behavioural evidence from navigation log files. *Journal of Computer Assisted Learning, 30*(3), 272-284.

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

Jovanovic, J., Pardo, A., Mirriahi, N., Dawson, S., & Gašević, D. (2017). An analytics-based framework to support teaching and learning in a flipped classroom. *Learning Analytics in the Classroom: Translating Learning Analytics Research for Teachers. Oxon: Routledge.*

Kallio, H., Pietilä, A. M., Johnson, M., & Kangasniemi, M. (2016). Systematic methodological review: developing a framework for a qualitative semi-structured interview guide. *Journal of Advanced Nursing, 72*(12), 2954-2965.

Kelly, A. E., Lesh, R. A., & Baek, J. Y. (2014). *Handbook of design research methods in education: Innovations in science, technology, engineering, and mathematics learning and teaching*: Routledge.

Kiefer, P., Giannopoulos, I., Raubal, M., & Duchowski, A. (2017). Eye tracking for spatial research: Cognition, computation, challenges. *Spatial Cognition & Computation, 17*(1-2), 1-19. doi: 10.1080/13875868.2016.1254634

Kinnebrew, J. S., Loretz, K. M., & Biswas, G. (2013). A contextualized, differential sequence mining method to derive students' learning behavior patterns. *JEDM| Journal of Educational Data Mining, 5*(1), 190-219.

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

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.

Köck, M., & Paramythis, A. (2011). Activity sequence modelling and dynamic clustering for personalized e-learning. *User Modeling and User-Adapted Interaction, 21*(1-2), 51-97. doi: 10.1007/s11257-010-9087-z

Kuhn, T. S. (2012). *The structure of scientific revolutions*: University of Chicago press.

Kumar, V., Venkatesan, R., & Reinartz, W. (2004). A purchase sequence analysis framework for targeting products, customers and time period. *Forthcoming in Journal of Marketing*.

Kurki, K., Järvenoja, H., Järvelä, S., & Mykkänen, A. (2017). Young children's use of emotion and behaviour regulation strategies in socio-emotionally challenging day-care situations. *Early Childhood Research Quarterly, 41*, 50-62.

Lan, M., & Lu, J. (2017). *Assessing the Effectiveness of Self-Regulated Learning in MOOCs Using Macro-level Behavioural Sequence Data.* Paper presented at the EMOOCs-WIP.

Lazakidou, G., & Retalis, S. (2010). Using computer supported collaborative learning strategies for helping students acquire self-regulated problem-solving skills in mathematics. *Computers & Education, 54*(1), 3-13.

Lin, B., Coburn, S. S., & Eisenberg, N. (2016). Self-regulation and reading achievement. *The cognitive development of reading and reading comprehension*, 67-86.

Liu, Z., Dev, H., Dontcheva, M., & Hoffman, M. (2016). *Mining, pruning and visualizing frequent patterns for temporal event sequence analysis.* Paper presented at the Proceedings of the IEEE VIS 2016 Workshop on Temporal & Sequential Event Analysis.

Lorigo, L., Haridasan, M., Brynjarsdóttir, H., Xia, L., Joachims, T., Gay, G., . . . Pan, B. (2008). Eye tracking and online search: Lessons learned and challenges ahead. *Journal of the Association for Information Science and Technology, 59*(7), 1041-1052. doi: 10.1002/asi.20794

Lubahn, D. B., Joseph, D. R., Sar, M., Tan, J.-a., Higgs, H. N., Larson, R. E., . . . Wilson, E. M. (1988). The human androgen receptor: complementary deoxyribonucleic acid cloning, sequence analysis and gene expression in prostate. *Molecular Endocrinology, 2*(12), 1265-1275. doi: 10.1210/mend-2-12-1265

Lupia, A., & Alter, G. (2014). Data access and research transparency in the quantitative tradition. *PS: Political Science & Politics, 47*(1), 54-59. doi: 10.1017/S1049096513001728

Lust, G. (2012). Opening the Black Box. Students' Tool-use within a Technology-Enhanced learning environment: An Ecological-Valid Approach.

Lust, G., Vandewaetere, M., Ceulemans, E., Elen, J., & Clarebout, G. (2011). Tool-use in a blended undergraduate course: In Search of user profiles. *Computers & Education, 57*(3), 2135-2144. doi: 10.1016/j.compedu.2011.05.010

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

Mannila, H., Toivonen, H., & Verkamo, A. I. (1997). Discovery of frequent episodes in event sequences. *Data mining and knowledge discovery, 1*(3), 259-289. doi: 10.1023/A:1009748302351

Martin, S. E., Shabanowitz, J., Hunt, D. F., & Marto, J. A. (2000). Subfemtomole MS and MS/MS peptide sequence analysis using nano-HPLC micro-ESI fourier transform ion cyclotron resonance mass spectrometry. *Analytical Chemistry, 72*(18), 4266-4274. doi: 10.1021/ac000497v

Masseglia, F., Teisseire, M., & Poncelet, P. (2002). *Real time Web usage mining with a distributed navigation analysis.* Paper presented at the Research Issues in Data Engineering: Engineering E-Commerce/E-Business Systems, 2002. RIDE-2EC 2002. Proceedings. Twelfth International Workshop on.

Mazon, J. M., Rossi, J. D., & Toledo, J. (2014). An optimal matching problem for the Euclidean distance. *SIAM Journal on Mathematical Analysis, 46*(1), 233-255. doi: 10.1137/120901465

Meredith, J. (1993). Theory building through conceptual methods. *International Journal of Operations & Production Management*, *13*(5), 3-11.

Mertens, D. M. (2014). *Research and evaluation in education and psychology: Integrating diversity with quantitative, qualitative, and mixed methods*: Sage publications.

Mills, M., Van de Bunt, G. G., & De Bruijn, J. (2006). Comparative research: Persistent problems and promising solutions. *International Sociology, 21*(5), 619-631. doi: 10.1177/0268580906067833

Molenaar, I., & Chiu, M. M. (2015). *Effects of sequences of socially regulated learning on group performance.* Paper presented at the Proceedings of the Fifth International Conference on Learning Analytics And Knowledge.

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

Moravcsik, A. (2014). Transparency: The revolution in qualitative research. *PS: Political Science & Politics, 47*(1), 48-53. doi: 10.1017/S1049096513001789

Müller, H., Naumann, F., & Freytag, J.-C. (2003). Data quality in genome databases.

Müller, N. S., Studer, M., Gabadinho, A., & Ritschard, G. (2010). *Analyse de séquences d'événements avec TraMineR.* Paper presented at the EGC.

Oliver, M., & Trigwell, K. (2005). Can 'blended learning' be redeemed. *E-learning, 2*(1), 17–26. doi: 10.2304/elea.2005.2.1.17

Paans, C., Molenaar, I., Segers, E., & Verhoeven, L. (2018). Temporal variation in children's self-regulated hypermedia learning. *Computers in Human Behavior*.

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

Perer, A., & Wang, F. (2014). *Frequence: Interactive mining and visualization of temporal frequent event sequences.* Paper presented at the Proceedings of the 19th international conference on Intelligent User Interfaces.

Perez, S., Massey-Allard, J., Butler, D., Ives, J., Bonn, D., Yee, N., & Roll, I. (2017). *Identifying productive inquiry in virtual labs using sequence mining.* Paper presented at the International Conference on Artificial Intelligence in Education.

Peterson, R. A. (2005). Problems in comparative research: The example of omnivorousness. *poetics, 33*(5), 257-282. doi: 10.1016/j.poetic.2005.10.002

Pintrich, P. R. (2004). A conceptual framework for assessing motivation and self-regulated learning in college students. *Educational Psychology Review, 16*(4), 385-407. doi: 10.1007/s10648-004-0006-x

Poole, M. S., Lambert, N., Murase, T., Asencio, R., & McDonald, J. (2016). Sequential analysis of processes. *The Sage handbook of process organization studies*, 254.

Pressley, M., Levin, J. R., & McDaniel, M. A. (1987). Remembering versus inferring what a word means: Mnemonic and contextual approaches.

Prinzie, A., & Van den Poel, D. (2007). Predicting home-appliance acquisition sequences: Markov/Markov for discrimination and survival analysis for modeling sequential information in NPTB models. *Decision Support Systems, 44*(1), 28-45. doi: 10.1016/j.dss.2007.02.008

Proctor, E. K., Powell, B. J., & McMillen, J. C. (2013). Implementation strategies: recommendations for specifying and reporting. *Implementation Science, 8*(1), 139. doi: 10.1186/1748-5908-8-139

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

Rahm, E., & Do, H. H. (2000). Data cleaning: Problems and current approaches. *IEEE Data Eng. Bull., 23*(4), 3-13.

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.

Roll, I., & Winne, P. H. (2015). Understanding, evaluating, and supporting self-regulated learning using learning analytics. *Journal of Learning Analytics, 2*(1), 7-12.

Roth, A., Ogrin, S., & Schmitz, B. (2016). Assessing self-regulated learning in higher education: a systematic literature review of self-report instruments. *Educational Assessment, Evaluation and Accountability, 28*(3), 225-250. doi: 10.1007/s11092-015-9229-2

Rotter, J. B. (1954). Social learning and clinical psychology.

Schnaubert, L., Heimbuch, S., & Bodemer, D. (2016). *Extracting selection strategies: comapring measures to analyze sequential data.* Paper presented at the EARLI SIG27 Measuring Learning Online, Oulu, Finland.

Schraw, G., Crippen, K. J., & Hartley, K. (2006). Promoting self-regulation in science education: Metacognition as part of a broader perspective on learning. *Research in Science Education, 36*(1-2), 111-139. doi: 10.1007/s11165-005-3917-8

Schraw, G., & Moshman, D. (1995). Metacognitive theories. *Educational Psychology Review, 7*(4), 351-371. doi: 10.1007/Bf02212307

Segedy, J. R., & Biswas, G. (2015). *Towards Using Coherence Analysis to Scaffold Students in Open-Ended Learning Environments.* Paper presented at the AIED Workshops.

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), 13-48.

Siadaty, M., Gašević, D., & Hatala, M. (2016). Associations between technological scaffolding and micro-level processes of self-regulated learning: A workplace study. *Computers in Human Behavior, 55*, 1007-1019. doi: 10.1016/j.chb.2015.10.035

Singleton, R., Straits, B. C., & Straits, M. (1999). Approaches to Social Research Oxford University Press. *New York and Oxford.*

Slater, S., Joksimović, S., Kovanovic, V., Baker, R. S., & Gasevic, D. (2017). Tools for educational data mining: A review. *Journal of Educational and Behavioral Statistics, 42*(1), 85-106. doi: 10.3102/1076998616666808

Sonnenberg, C., & Bannert, M. (2015). Discovering the effects of metacognitive prompts on the sequential structure of SRL-processes using process mining techniques. *Journal of Learning Analytics, 2*(1), 72-100.

Srikant, R., & Agrawal, R. (1996). *Mining sequential patterns: Generalizations and performance improvements.* Paper presented at the International Conference on Extending Database Technology.

Stackebrandt, E., & Goebel, B. (1994). Taxonomic note: a place for DNA-DNA reassociation and 16S rRNA sequence analysis in the present species definition in bacteriology. *International Journal of Systematic and Evolutionary Microbiology, 44*(4), 846-849.

Stark, D., & Vedres, B. (2012). Social Sequence Analysis. *The Emergence of Organizations and Markets*, 347-364.

Studer, M., Mueller, N. S., Ritschard, G., & Gabadinho, A. (2010). *Classer, discriminer et visualiser des séquences d'événements.* Paper presented at the EGC.

Taub, M., Azevedo, R., Bouchet, F., & Khosravifar, B. (2014). Can the use of cognitive and metacognitive self-regulated learning strategies be predicted by learners' levels of prior knowledge in hypermedia-learning environments? *Computers in Human Behavior, 39*, 356-367.

Taub, M., Azevedo, R., Bradbury, A. E., Millar, G. C., & Lester, J. (2017). Using sequence mining to reveal the efficiency in scientific reasoning during STEM learning with a game-based learning environment. *Learning and Instruction.*

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.

Taub, M., Mudrick, N. V., Azevedo, R., Millar, G. C., Rowe, J., & Lester, J. (2016). *Using Multi-level Modeling with Eye-Tracking Data to Predict Metacognitive Monitoring and Self-regulated Learning with C rystal I sland.* Paper presented at the International Conference on Intelligent Tutoring Systems.

Tsai, C.-W., Shen, P.-D., & Tsai, M.-C. (2011). Developing an appropriate design of blended learning with web-enabled self-regulated learning to enhance students' learning and thoughts regarding online learning. *Behaviour & Information Technology, 30*(2), 261-271. doi: 10.1080/0144929x.2010.514359

Van der Aalst, W. M. (2016). *Process mining: data science in action*: Springer.

Van der Merwe, W. A. J. (2013). *Towards a conceptual model of the relationship between corporate trust and corporate reputation.* University of Pretoria.

Van Krevelen, D. W., & Te Nijenhuis, K. (2009). *Properties of polymers: their correlation with chemical structure; their numerical estimation and prediction from additive group contributions*: Elsevier.

Van Laer, S., & Elen, J. (2016). Adults' Self-Regulatory Behaviour Profiles in Blended Learning Environments and Their Implications for Design. *Technology, Knowledge and Learning*, 1-31.

Van Laer, S., & Elen, J. (2018). An Instrumentalized Framework for Supporting Learners' Self-regulation in Blended Learning Environments. In M. J. Spector, B. B. Lockee, & M. D. Childress (Eds.), *Learning, Design, and Technology: An International Compendium of Theory, Research, Practice, and Policy*: Springer, Cham.

Van Laer, S., Jiang, L., & Elen, J. (2018). The Effect of Cues for Reflection on Learners' Self-regulated Learning through Changes in Learners' Learning Behaviour and Outcomes. *Computers & Education*, under review.

Veenman, M. V. (2007). The assessment and instruction of self-regulation in computer-based environments: a discussion. *Metacognition and Learning, 2*(2), 177-183.

Veenman, M. V., & Alexander, P. (2011). Learning to self-monitor and self-regulate. *Handbook of research on learning and instruction*, 197-218.

Veenman, M. V., Bavelaar, L., De Wolf, L., & Van Haaren, M. G. (2014). The on-line assessment of metacognitive skills in a computerized learning environment. *Learning and Individual Differences, 29*, 123-130. doi: 10.1016/j.lindif.2013.01.003

Veenman, M. V., Prins, F. J., & Verheij, J. (2003). Learning styles: Self-reports versus thinking-aloud measures. *British Journal of Educational Psychology, 73*(3), 357-372.

Veenman, M. V., Van Hout-Wolters, B. H., & Afflerbach, P. (2006). Metacognition and learning: Conceptual and methodological considerations. *Metacognition and Learning, 1*(1), 3-14.

Williamson, G. (2015). Self-regulated learning: an overview of metacognition, motivation and behaviour.

Winne, P. (2016). Self-regulated learning. *SFU Educational Review, 1*(1).

Winne, P. H. (2005). Key Issues in modeling and applying research on self-regulated learning. *Applied Psychology, 54*(2), 232-238.

Winne, P. H. (2010). Improving measurements of self-regulated learning. *Educational Psychologist, 45*(4), 267-276.

Winne, P. H. (2014). Issues in researching self-regulated learning as patterns of events. *Metacognition and Learning, 9*(2), 229-237. doi: 10.1007/s11409-014-9113-3

Winne, P. H. (2018). Theorizing and researching levels of processing in self-regulated learning. *British Journal of Educational Psychology, 88*(1), 9-20.

Winne, P. H., & Baker, R. S. (2013). The potentials of educational data mining for researching metacognition, motivation and self-regulated learning. *JEDM| Journal of Educational Data Mining, 5*(1), 1-8.

Winne, P. H., & Hadwin, A. F. (1998). Studying as self-regulated learning. *Metacognition in educational theory and practice, 93*, 27–30.

Winne, P. H., Nesbit, J. C., & Popowich, F. (2017). nStudy: A System for Researching Information Problem Solving. *Technology, Knowledge and Learning, 22*(3), 369-376. doi: 10.1007/s10758-017-9327-y

Winne, P. H., & Perry, N. E. (2000). Measuring self-regulated learning.

Yang, X., Li, J., & Xing, B. (2018). Behavioral patterns of knowledge construction in online cooperative translation activities. *The Internet and Higher Education, 36*, 13-21. doi: 10.1016/j.iheduc.2017.08.003

Zaki, M. J. (2001). SPADE: An efficient algorithm for mining frequent sequences. *Machine learning, 42*(1-2), 31-60. doi: 10.1023/A:1007652502315

Zhou, M. (2016). Data pre-processing of student e-learning logs *Information Science and Applications (ICISA) 2016* (pp. 1007-1012): Springer.

Zhou, M., Xu, Y., Nesbit, J. C., & Winne, P. H. (2010). Sequential pattern analysis of learning logs: Methodology and applications. *Handbook of educational data mining, 107*, 107-121.

Zimmerman, B. J. (1986). Becoming a self-regulated learner: Which are the key subprocesses? *Contemporary Educational Psychology, 11*, 307–313.

Zimmerman, B. J. (1990). Self-regulated learning and academic achievement: An overview. *Educational Psychologist, 25*(1), 3–17. doi: 10.1207/s15326985ep2501_2

Zimmerman, B. J. (1998). Academic studing and the development of personal skill: A self-regulatory perspective. *Educational Psychologist, 33*, 73–86.

Zimmerman, B. J. (2000). Self-Efficacy: An Essential Motive to Learn. *Contemporary Educational Psychololgy, 25*(1), 82-91. doi: 10.1006/ceps.1999.1016

Zimmerman, B. J., & Pons, M. M. (1986). Development of a Structured Interview for Assessing Student Use of Self-Regulated Learning Strategies. *American Educational Research Journal, 23*(4), 614-628. doi: 10.3102/00028312023004614

Zimmerman, B. J., & Schunk, D. H. (2001). *Self-regulated learning and academic achievement: Theoretical perspectives*: Routledge.