Received: 27 May 2019 | Accepted: 27 March 2022

DOI: 10.1111/jcal.12675

ARTICLE

Journal of Computer Assisted Learning WILEY

Exploring the link between self-regulated learning and learner behaviour in a massive open online course

Renée S. Jansen | Anouschka van Leeuwen | Jeroen Janssen | Liesbeth Kester

Utrecht University, Department of Education, Utrecht, The Netherlands

Correspondence
Renée S. Jansen, Department of Education, Utrecht University, Heidelberglaan 1, 3584 CS Utrecht, The Netherlands.
Email: R.S.Jansen-14@umcutrecht.nl

Funding information
Nederlandse Organisatie voor Wetenschappelijk Onderzoek, Grant/Award Number: 405-15-705

Abstract

Background: Learners in Massive Open Online Courses (MOOCs) are presented with great autonomy over their learning process. Learners must engage in self-regulated learning (SRL) to handle this autonomy. It is assumed that learners' SRL, through monitoring and control, influences learners' behaviour within the MOOC environment (e.g., watching videos). The exact relationship between SRL and learner behaviour has however not been investigated.

Objectives: We explored whether differences in SRL are related to differences in learner behaviour in a MOOC. As insight in this relationship could improve our understanding of the influence of SRL on behaviour, could help explain the variety in online learner behaviour, and could be useful for the development of successful SRL support for learners.

Methods: MOOC learners were grouped based on their self-reported SRL. Next, we used process mining to create process models of learners' activities. These process models were compared between groups of learners.

Results and conclusions: Four clusters emerged: average regulators, help seekers, self-regulators, and weak regulators. Learners in all clusters closely followed the designed course structure. However, the process models also showed differences which could be linked to differences in the SRL scores between clusters.

Takeaways: The study shows that SRL may explain part of the variability in online learner behaviour. Implications for the design of SRL interventions include the necessity to integrate support for weak regulators in the course structure.

KEYWORDS
learner behaviour, MOOC, online education, process mining, SRL

1 | INTRODUCTION

Learners in a massive open online course (MOOC) experience much more autonomy over their learning process compared to learners in traditional campus-based education (Wang et al., 2013). Learners can study at any time, any place, and any pace they prefer, since course materials are available online over longer periods of time, and they can be studied by MOOC participants without guidance of a teacher. To handle the autonomy offered to them, students must engage in self-regulated learning (SRL) in MOOCs (Azevedo & Aleven, 2013; Beishuizen & Steffens, 2011; Garrison, 2003; Kizilcec et al., 2017; Kizilcec & Halawa, 2015; Wang et al., 2013; Waschull, 2001). To learn successfully in a MOOC, learners must take control of their own learning process. MOOC learners that are unable to adequately self-regulate their learning are likely to drop out (Hew & Cheung, 2014; Kizilcec & Halawa, 2015).

Self-regulated learners are actively involved in their learning, and they make conscious decisions about what, where, and how they study (Zimmerman, 2002). It involves activities such as planning, monitoring, time management, and help seeking. Nelson and Narens (1990) described SRL as a continuous cycle between monitoring and control (see Figure 1). Learners engage in learning activities to perform a task. These activities are overt; they can be observed by others. While working on the task, learners monitor their progress. As a result, learners form a metacognitive representation of their learning at a *meta-level*. Based on the progress monitored, and the gap between current and desired performance, learners control their overt learning activities at the *object-level*. Monitoring and control, which are covert activities, thereby help self-regulated learners to adapt their learning activities to the task at hand (Littlejohn et al., 2016).

The automatic storage of all learners' activities in a MOOC learning environment into trace data enables researchers to study the relationship between covert SRL and overt learner behaviour at a level of detail that is not feasible in traditional education. In trace data, all learner behaviour is stored at a very fine granularity over the time span of the whole course. Empirical data of this kind cannot be collected in traditional education. In this study, we will make use of the opportunities trace data offer to study the relationship between SRL and learner behaviour in a MOOC.

Investigating the relationship between SRL and learner behaviour in a MOOC is however not only interesting for the empirical data that it provides. It also improves our understanding of the influence of SRL on online course behaviour. Due to the autonomy provided to them, MOOC learners can study in highly varying ways (Kizilcec et al., 2013). Research has shown that learners indeed make use of this opportunity and found great variety in the way learners study in online education: learners for instance differ in terms of the amount of material they complete, the (order of) activities they engage in, but also in the their forum activities, and in the timing of and time between their learning sessions (e.g., Goda et al., 2015; Jovanović et al., 2017; Kizilcec et al., 2013; Kovanović et al., 2015; Maldonado-Mahauad et al., 2018; Saint et al., 2018). Theoretically however, little is known about the origin of the variety in learner behaviour (Li & Baker, 2018). Differences between learners concerning, for example, prior knowledge and SRL, may be the cause of these differences in course behaviour (Li & Baker, 2018). More research on how differences between learners influence learner behaviour is necessary (Deng et al., 2019). The importance of SRL for successful learning in MOOCs leads us to focus on the relationship between SRL and learner behaviour in this study.

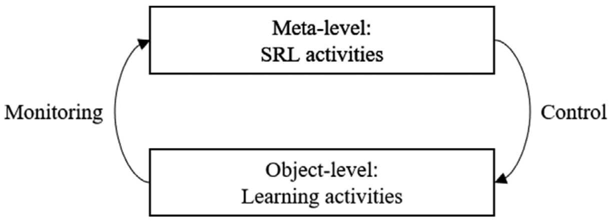

FIGURE 1 The relationship between monitoring and control (Nelson & Narens, 1990)

Since insufficient SRL can lead to student dropout, multiple researchers have attempted to support learners' SRL by implementing an SRL intervention in a MOOC (Davis et al., 2018; Kizilcec et al., 2016; Yeomans & Reich, 2017). Exploring the influence of SRL on learner behaviour could help increase the impact of such SRL interventions. While compliance with the SRL support offered in these studies increased both learners' course activity as well as their course completion, these interventions suffered from low compliance by learners: many learners did not engage with the SRL support offered. It is known that weak learners often find it most challenging to identify their support needs (Clarebout et al., 2010; Clarebout & Elen, 2006). It is therefore likely that learners who needed help most, did not engage with the SRL support. Increased knowledge of how learners' SRL influences how learners behave in the MOOC environment, especially of how weak self-regulating students behave, may help identify ways in which support could be implemented to increase learner compliance. Exploration of the impact of learners' SRL on their learning process in MOOCs may thereby help determine how SRL support should best be implemented.

SRL can thus be considered important for student learning in MOOCs, and there is considerable theoretical and practical value in investigating the relationship between SRL and learner behaviour. This would provide data on the relationship between SRL and learner behaviour, help explain the variability in online learner behaviour and assist in the implementation of SRL support in MOOCs. Nevertheless, research on the influence of learners' SRL on learner behaviour within MOOCs is limited. In the section below, we present existing research on the relationship between SRL and learner behaviour in MOOCs and describe how the current study extends this knowledge.

1.1 | Literature review

One of the first studies to link learners' activities, captured in trace data, with learners' SRL was conducted by Hadwin et al. (2007). For eight learners, the association between specific self-reported SRL (measured by means of a questionnaire) and learners' trace data in a single study session was analysed. Trace data provided additional, and in some cases conflicting, information to learners' self-reported SRL. While the authors mostly focused on single questionnaire items and absolute frequencies of learners' activities, their results already showed that a better understanding of students' SRL could be gained by adding trace data to questionnaire data (Hadwin et al., 2007; Winne, 2010; Zimmerman, 2008). In a more recent study, Kizilcec et al. (2017) also investigated the relationship between learners' self-reported SRL and their learner behaviour as measured with trace data. In contrast to the study conducted by Hadwin et al. (2007), Kizilcec et al. (2017) focused on scores on SRL scales, instead of on individual items, and on the frequency of transitions from one activity to the next, instead of on absolute frequencies of activities. For instance, they explored the relation between goal setting (an individual scale from the employed SRL questionnaire) and the action of revisiting a lecture after watching a lecture. Overall, they found that learners who reported more engagement in SRL activities were more likely to revisit materials they had already completed compared to learners who reported less engagement in SRL activities. The results thereby showed that SRL and learner behaviour are related. This finding indicates that the relationship between SRL and learner behaviour is not only found when analysing questionnaire data per item (Hadwin et al., 2007), but also when analysing questionnaire data at a higher level of aggregation, namely per scale (Kizilcec et al., 2017)

The approach taken by Kizilcec et al. (2017) however still provides limited insight in the influence of SRL on learner behaviour. The six SRL scales present in the questionnaire (e.g., strategic planning, help seeking) were all individually correlated to the 36 behavioural transitions that were studied. Learners' scores on SRL components are, however, related (Sitzmann & Ely, 2011). Sitzmann and Ely (2011) for instance found metacognition to be correlated to time management and help seeking. It is therefore likely that the SRL scales were also correlated in the research conducted by Kizilcec et al. (2017). Multiple scales correlated to the same transition, namely revisiting assessments after passing an assessment. The correlation between SRL scales may explain why most of the scales were found to be significantly correlated to that same behavioural transition. In the current study, we take the correlation between aspects of SRL into account by studying SRL as a single construct by clustering learners.

Furthermore, by analysing learning as a collection of individual transitions, the approach taken by Kizilcec et al. (2017) ignored the presence of a course structure that (partly) determined students' learning process (Bannert et al., 2014). The ordering of learners' activities is governed by this structure. In the present study, learners for instance transitioned from content videos to self-test questions when they followed the designed order of learning activities. The influence of the course structure is neglected when analysing individual transitions but can be incorporated when analysing learner processes (Bannert et al., 2014). Moreover, analysis of students' activities as a process instead of as individual transitions also presents a better representation of students' learning. As learning is cumulative, activities inherently build upon each other (Reimann, 2009). Larger sequences of learning than individual transitions should thus be taken into account to accurately model students' learning process. We therefore analyse learners' activities in the MOOC through process mining. Process mining allows for the analysis of large samples of ordered (i.e. time-stamped) activity data (Sonnenberg & Bannert, 2015, 2018). We thereby analyse all transitions at once, instead of focusing on each transition separately.

We know of only a single study in which online learning processes have been linked to SRL. Maldonado-Mahauad et al. (2018) focused on the activities learners engaged in between starting a learning session and ending a learning session. Process mining was used to find all the different sequences of activities. Each session was then classified based on the overall occurring activity. Six types of learning sessions emerged, including only watching video lectures and attempting an assessment followed by watching the accompanying video lecture. For each type of learning session, the authors provided an explanation in terms of SRL that might underlie the learning activities performed in that session. For example, they suggested that the sequence of watching a video lecture followed by completing an assessment might signal the use of the SRL strategy self-evaluation. However, SRL was not measured in this study and the potential SRL explanations of the learning sessions are thus not based on data but on interpretation by the authors. In the current study we also focus on the relationship between SRL and the order of learning activities. In contrast to Maldonado-Mahauad et al. (2018) we combine learners' trace data with learners' SRL measured with a questionnaire.

1.2 | The current study

In the current study, we explore the relationship between SRL and learner behaviour in a MOOC. Due to the autonomy offered to learners in MOOCs, SRL is of considerable importance for successful MOOC learning (e.g., Kizilcec & Halawa, 2015). Insight into the relationship between SRL and learner behaviour has both practical as well as theoretical relevance, as it helps determine how SRL support can best be implemented in MOOCs and improves our understanding of how SRL influences learner behaviour. Learners' SRL will be measured with a questionnaire (Jansen et al., 2017). The trace data captured in the MOOC learning environment will be used to access learners' behaviour (Hadwin et al., 2007; Kizilcec et al., 2017; Maldonado-Mahauad et al., 2018). The relationship between learners' SRL and their learner behaviour will be analysed by first clustering learners into groups with similar SRL and then analysing the order of their learning activities with process mining (Bannert et al., 2014; Maldonado-Mahauad et al., 2018). We hereby extend existing research in two ways. We analyse SRL as a construct instead of as separate, independent scales, and we analyse behaviour processes instead of individual transitions.

2 | METHOD

2.1 | Context

Data were collected in a MOOC on Environmental Sustainability offered by Wageningen University, The Netherlands, on the online learning platform edX. The MOOC ran from September 2016 to November 2016 and consisted of seven modules. The first module was an introductory module, called module 0, and contained the course manual and introductory videos of the lecturers. Module 1–6 were all content modules. Each consisted of an introductory video, approximately four content videos each with one or two recap (i.e., self-test) questions, a summary video, a practice test, and a graded test. All questions in the course were multiple choice questions. Module 6 was the final module, which contained both a graded test and the final exam. The exam consisted of writing a peer-assessed essay. A course forum was connected to the course environment for the course instructors and designers and the course participants. Browsing and posting on the forum were not required in the course, but the forum could be accessed at any time. The study pace advised by the course designers was one module per week, but learners were free to study at a faster or slower pace.

2.2 | Participants

MOOC learners were presented a questionnaire which could be answered voluntarily and anonymously focused on their SRL. While there were more learners in this MOOC, we focus in this study on the learners who answered the questionnaire ($n = 73$). All participants who answered all questions identically were removed, as they were considered outliers due to the lack of deviation in their answers ($n = 4$). The remaining participants formed the sample of the present study ($n = 69$). Their mean age was 38.8, 40.6% were male.

2.3 | Measurements

2.3.1 | SRL

SRL was measured with the self-regulated online learning questionnaire (SOL-Q; Jansen et al., 2017). This questionnaire consisted of 36 items and measured learners' SRL using five different scales: metacognitive skills (17 items, $\alpha = 0.90$), time management (3 items, $\alpha = 0.73$), environmental structuring (5 items, $\alpha = 0.73$), persistence (5 items, $\alpha = 0.69$), and help seeking (5 items, $\alpha = 0.89$). In the same order, example items of the five scales are 'I ask myself questions about what I am to study before I begin to learn for this online course', 'I find it hard to stick to a study schedule for this online course', 'I know where I can study most efficiently for this online course', 'When my mind begins to wander during a learning session for this online course, I make a special effort to keep concentrating' and 'When I am not sure about some material in this online course, I check with other people'.

All questions were answered on a 7-point Likert scale ranging from 'not at all true for me' to 'very true for me'. The questionnaire was incorporated in the course environment as a voluntary assignment at the end of module 2. At that point, learners were able to reflect on their SRL during the MOOC. Learners were stimulated to answer the questions based on their experiences in the online course instead of based on their experience with learning in general by including the phrase 'in this online course' in all questions.

2.3.2 | Learner behaviour

Learner behaviour was defined as learners' engagement in thirteen learning activities. These learning activities (see Figure 2) were derived from the course structure because these activities formed the main components of the MOOC. The learning activities were: watching introductory videos, content videos, and summary videos (1–3), answering multiple-choice recap questions correctly/incorrectly, practice questions correct/incorrect, and graded questions correct/incorrect (4–9), handing in the essay assignment (10), assessing peers (11), and browsing and posting on the forum (12–13). The order of these activities as intended by the course designers is displayed in Figure 2. We filtered information on the thirteen activities analysed from the trace data. In the trace data, all learner activities in the MOOC environment were automatically stored including a timestamp and a user ID.

When following the intended process, a learner would start each module by watching the introductory video. A learner would continue with watching the first content video and answering the one or two associated recap questions. As there were two questions for most videos, and questions could be re-answered, the learner could move between correctly and incorrectly answering recap questions. The learner would continue watching content videos and answering recap questions until all content videos included in the module had been viewed. The learner would then watch the summary video and make the practice test. The practice test consisted of multiple questions and therefore the learner could transition between answering practice questions correctly and incorrectly. There were no consequences for answering a recap or practice question incorrectly, and the correct answer was shown as soon as the question was answered incorrectly. Therefore, the learner would know the correct answer to a recap or practice test question also after answering the question incorrectly. After the practice test, the learner would work on the graded test. This test also consisted of multiple questions, making it possible for the learner to have an incorrect question follow a correct question or reversed. After answering the final question, either correctly or incorrectly, the learner would start working on the next module by watching the introductory video. After finishing the graded test of the

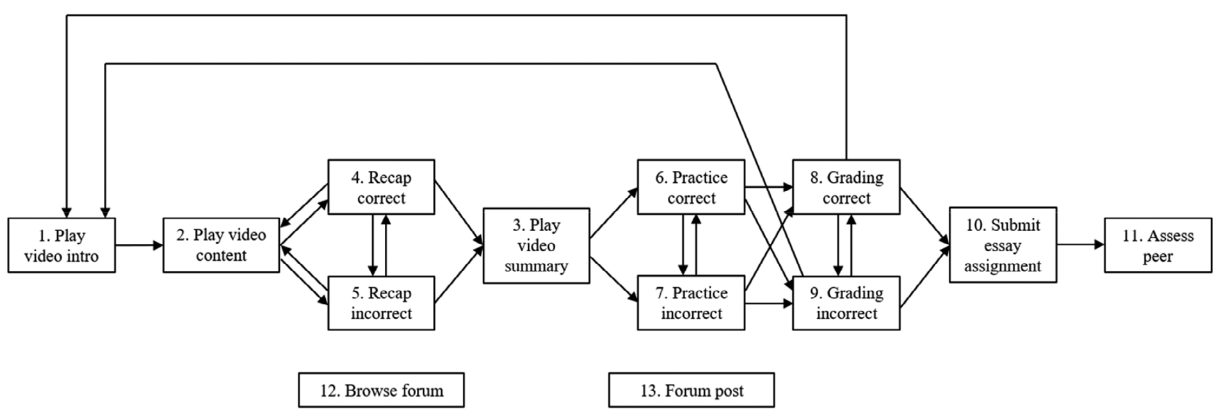

FIGURE 2 Process model of the course structure as intended by the course designers

sixth module, the learner would hand in a peer-assessed essay. After handing in the essay, the learner had to grade the work of at least four others, before the learner's own grades would become available. If the learner's own work was peer-assessed to be a pass, the learner completed the course after grading four others.

2.4 | Procedure

Learners could work with the course material in any order and at any pace they liked. The questionnaire on SRL was presented as a voluntary assignment at the end of module 2. Completion of the questionnaire took approximately 15 minutes. By completing the questionnaire on SRL, learners gave their informed consent and thereby gave permission to link their questionnaire responses to their trace data. The trace data were later retrieved from the edX server. As the current study focuses on the relation between interaction with course materials and reported SRL, only the trace data for those learners who filled out the SRL questionnaire were further analysed. Permission for this study was attained from the institution's ethics committee.

2.5 | Data analysis

In the current study, process models of groups with different self-reported SRL were compared to investigate how self-reported SRL is related to the order of learners' activities within the MOOC. As it is not feasible to compare the process models of all individual learners in the sample, learners first had to be clustered into groups with similar SRL before process models could be created.

2.5.1 | Cluster analysis

Procedures as outlined in Mooi and Sarstedt (2010) were followed to conduct cluster analysis with a small sample. The scale scores (i.e., mean score per scale) of the five scales in the SRL questionnaire were used as the basis for clustering. The first step in cluster analysis was the exclusion of outliers. They are not part of any cluster and can severely influence the cluster solution (Milligan, 1980). It is therefore advised to remove these cases before conducting cluster analyses by using the single linkage Euclidian distance algorithm (Mooi & Sarstedt, 2010). Seven cases were removed as outliers. Next, hierarchical cluster analysis was conducted with the remaining 62 cases using Ward's method (Mooi & Sarstedt, 2010). Cases are hereby segregated into clusters by combining cases that lead to the smallest increase in total variance per cluster. A four cluster solution led to the most equal distribution of learners over clusters and could also be best interpreted. The clusters were furthermore similar to the clusters found in previous studies in which learners were clustered based on their self-reported SRL (Barnard-Brak et al., 2010; Dörrenbächer & Perels, 2016; Ning & Downing, 2015). The four cluster solution was therefore selected as the final clustering. An overview of the clusters and the SRL scores of the learners within them can be found in Table 1.

The four clusters were labelled based on the reported SRL data. The first and largest cluster is a group of *average regulators*. Learners in this cluster reported average levels of SRL compared to the other clusters. The second group consists of *help seekers*. While learners in this group reported average levels of most SRL activities, it stands out that they indicated more engagement with help seeking behaviour than the other clusters of learners. This indicates that these learners were aware of other learners in the course and that they wanted to engage with them to improve their learning. The third cluster is formed by the *self-regulators*. Learners in this cluster indicated the highest level of metacognitive skills, environmental structuring, and persistence. Their level of self-reported time management was almost as high as that of the average regulators, who indicated the highest score (4.47 versus 4.50). The *self*-regulators only scored lower on the help seeking scale than the help seekers. This, therefore, is a cluster of learners who indicated high *self*-regulated learning. The fourth and final cluster are the *weak regulators*. This cluster is exemplified by learners with lower scores than all other groups on the five SRL scales. Learners in this cluster appear to engage in the course without a clear strategy and without planning their study behaviour.

2.5.2 | Process mining

After clustering the learners, process mining was used to analyse the trace data per cluster (Bannert et al., 2014; Maldonado-Mahauad et al., 2018). With process mining, process models are created to compare process data between individuals, or between groups of individuals. Thereby, process mining allows for the analysis of temporal patterns in the data. The typical transitions (i.e., edges) of learners between activities (i.e., nodes) within each cluster are visualized, while atypical, infrequent transitions are removed to handle noise in the trace data. We analysed the trace data that related to interactions focused on whole activities, such as watching a video. We did not zoom in on finer grained activities, such as pausing a video, or navigating between pages. The activities included are presented in Figure 2.

Process mining was conducted with ProM 6.6 and the fuzzy miner algorithm (see Bannert et al., 2014). The settings used for the fuzzy miner algorithm in the current study are similar to those used in the study conducted by Bannert et al. (2014). However, as we were interested in the transitions between the thirteen activities specified, no activities were removed from the models even if they appeared only very infrequently in the trace data; the node filter cutoff was set to 0 to retain all activities in the resulting models. Furthermore, Bannert et al. (2014) only retained the most significant and frequent relations (edge filter cutoff 0.200). We preferred a greater level of detail to result from process mining. We set the edge filter cutoff at 0.500 to retain more transitions in the model. Self-loops (i.e., one

TABLE 1 Descriptives of the self-regulated online learning questionnaire per cluster

	Average regulators ($n = 22$)	Help seekers ($n = 15$)	*Self*-regulators ($n = 10$)	Weak regulators ($n = 15$)
Metacognitive skills	4.44	4.95	5.52	3.93
Time management	4.50	4.38	4.47	2.62
Environmental structuring	5.32	5.01	6.42	4.99
Persistence	4.28	4.52	5.94	3.73
Help seeking	1.45	3.65	1.82	1.55

Note: All scales on a range from 1 to 7.

activity to the same activity) were present in the data, but were ignored while creating the process models. This was done as self-loops were so frequently occurring (switching between pause and play of a single video, answering multiple questions in a row), that they would make all other transitions too infrequent to be included in the process models. As we were interested in how learners' transition from one learning activity to another, self-loops were not considered when determining the importance of transitions for the process models.

3 | RESULTS

To analyse the relationship between learners' SRL and their behaviour, we compared the behaviour of learners within the clusters based on the thirteen learning activities specified in the method. Process models were created for each of the four clusters of learners with similar self-reported SRL. The four resulting process models can be found in Figures 3–6.

The process models showed that learners in all clusters generally followed the course activities in the order designed and intended by the course designers as presented in Figure 2; most of the transitions in Figure 2 were also visible in the four process models. In contrast, learners' engagement with the recap questions showed a deviation from the course structure in all clusters. In most cases, two recap questions were connected to a content video. The process models all showed that while learners may sometimes have answered the first recap question incorrectly after watching a content video, this transition was so infrequent that it was removed from the process models (activity 2 to activity 5). The most traversed path was from watching a content video, to answering a recap question correctly, to answering a recap question incorrectly (2–4–5). Transitions from recap question incorrect to recap question correct (5–4) were also observed and this could indicate students who corrected their wrong answer.

The transitions originating from incorrectly answered recap questions show the first major difference between clusters. Learners in all four clusters answered recap questions correctly before continuing with other learning activities (4–3). For learners in three clusters (i.e., average regulators, help seekers, and self-regulators) this is the only displayed transition after answering a recap question incorrectly. Learners in the weak regulators cluster however also had a frequently occurring path from incorrectly answered recap questions to watching a summary video (5–3). Learners in the average regulators and self-regulators clusters transitioned from answering a practice question incorrectly to answering a graded question correctly (7–8). As the correct answer to a recap or practice question was presented after answering the question incorrectly, transitioning from an incorrectly answered recap question or from an incorrectly answered practice question to a next activity, was therefore not better or worse than first answering the recap or practice question correctly before moving to the next activity.

The process models of several clusters also showed skipping of steps intended by the course designers. For average regulators and help seekers, transitioning from a content video to the practice test, thereby skipping the summary video was a frequent alternative (2–6). For average regulators, self-regulators, and weak regulators immediately answering a recap question after watching the introductory video was a frequent alternative to watching the content video first (1–4). While possible in the course design, none of the process models showed direct skipping of the introductory video, as there were no direct relations from answering the graded questions to watching the content videos (8–2 or 9–2). All process models did include transitions from answering graded questions correctly, to browsing the forum, back to watching content videos (8–12–2). Indirect skipping of the introductory video may thus have occurred.

When comparing the process models with the process intended by the course designers, it should also be noted that only for the self-regulators the transition from answering a graded test question to submitting the essay assignment was present (8–10). When following the order of the online course as designed, this was the order in which one should arrive at the final assignment. In all process models however, other transitions to this activity were present. All process models included a transition from watching a summary video to submitting the essay (3–10).

It was not mandatory for learners to browse or post on the forum in order to follow or finish the course, although engagement with other learners through the forum could be helpful. When we zoom in on learners' forum interactions, we first analysed the help seekers. While help seekers only engaged in forum interactions after answering a graded test, and not at other moments in time, this transition is their only transition back from answering graded test questions (8–12–2). Learners in all other clusters transitioned directly from answering graded questions to watching an introductory video, likely of the

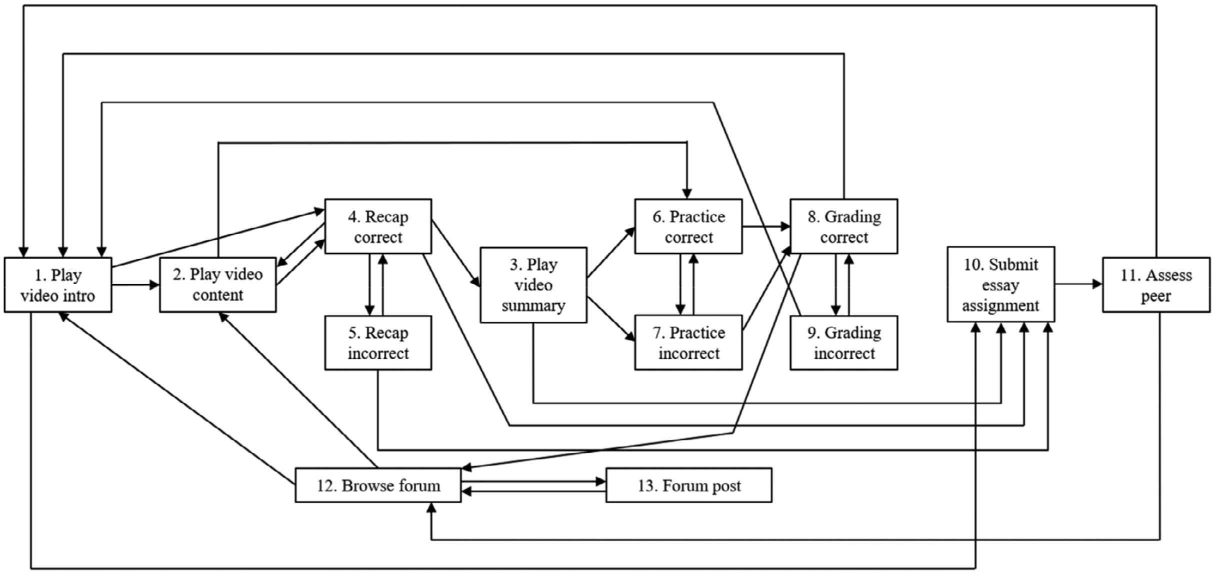

FIGURE 3 Process model for the cluster of *average regulators*

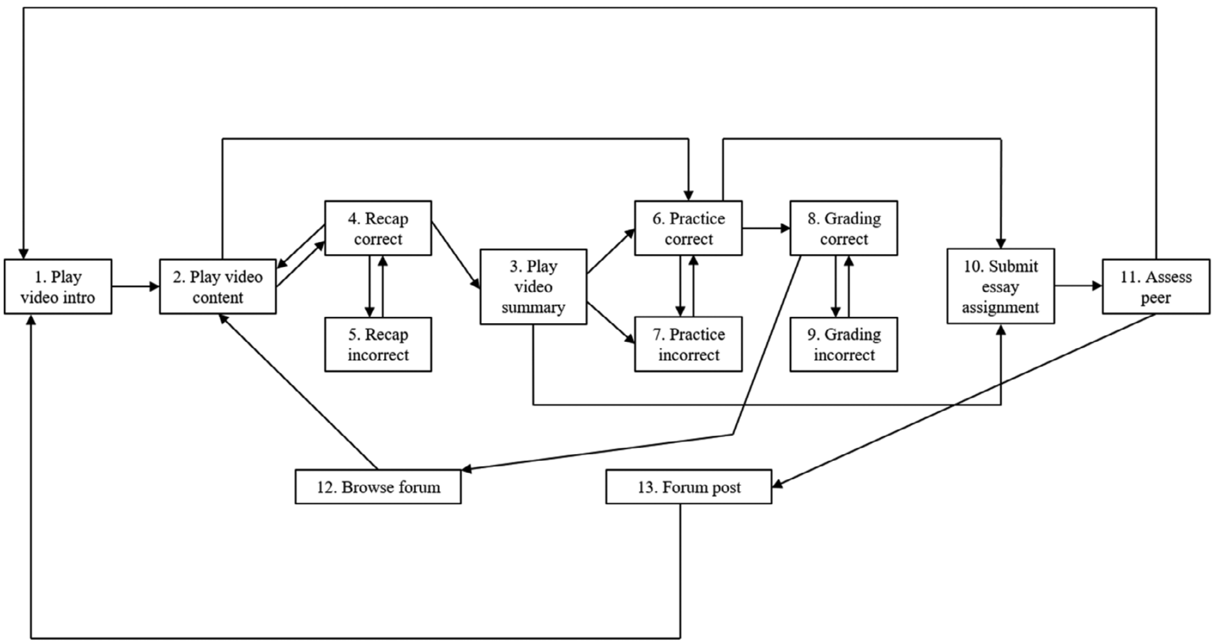

FIGURE 4 Process model for the cluster of *help seekers*

next module (8–1 or 9–1). Furthermore, for all other clusters, posting on the forum solely occurred in response to browsing the forum. Learners in the help seekers cluster were the only group that had no transition between browsing and posting. For the self-regulators, browsing and posting on the forum were highly integrated in their learning process; their process model showed a large number of transitions from and to browsing and posting. For the average regulators and the weak regulators, forum activities were present, but these were less integrated in their learning process.

To sum up, two statements can be made concerning how the different clusters of learners interacted with the course materials. First, learners in all clusters generally followed the course in the order intended by the course designers. The intended course structure is visible in all four process models. This also explains why the process models of the four different clusters show similarities. Second, however, there were also differences between the process models of the clusters. The clearest differences occurred considering skipping of activities, and browsing and posting on the forum.

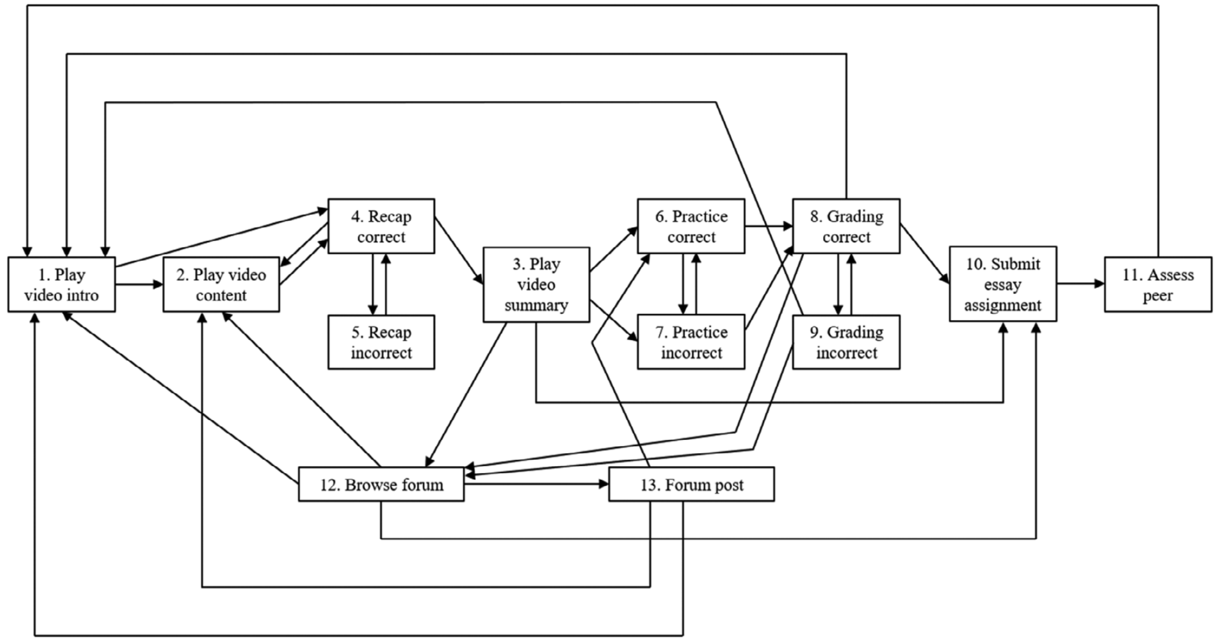

FIGURE 5 Process model for the cluster of *self-regulators*

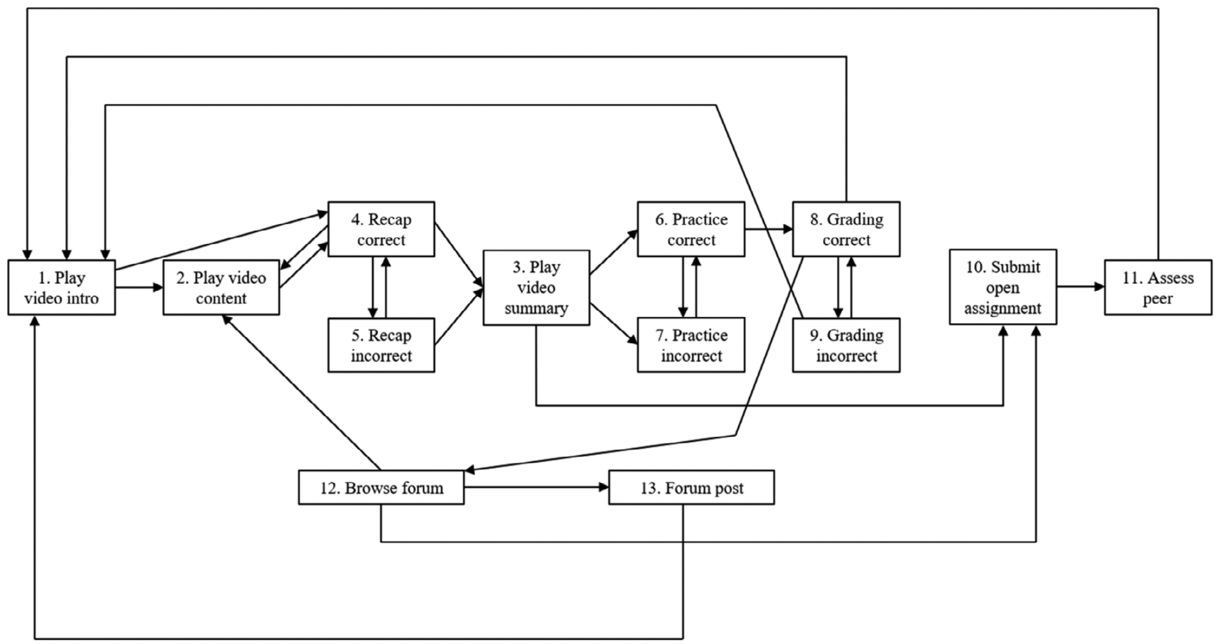

FIGURE 6 Process model for the cluster of *weak regulators*

4 | DISCUSSION

In the current study, we explored the relationship between SRL and learner behaviour in a MOOC. Process mining was used to compare learning processes. In order to conduct process mining, learners were first clustered based on their self-reported SRL. Four clusters emerged: average regulators, help seekers, self-regulators, and weak regulators. Next, the behaviour of learners in the different clusters was compared by using the trace data stored in the online course environment. Specifically, we looked at their learning processes in terms of thirteen learning activities (Figure 2).

Two general conclusions could be drawn based on the comparison between the process models of the four clusters. First, the process models showed similarities between clusters. Learners in all clusters were guided by the course structure implemented by the course designers, and this intended course structure was visible in all four process models. The MOOC from which the learner data was analysed in this study had a clear structure. Learners were guided in their learning process as all modules followed the same sequence and incorporated introduction and summary videos. Thereby the course design likely reduced the need for learners to regulate their learning.

However, SRL remained important as learners were still free to study what, where and when they wanted. The need for self-regulation is in line with the second general conclusion: differences between clusters were present, and these differences in process models could be interpreted in light of differences in SRL scores. The average regulators showed the greatest variety in the transitions present in their process model. Their SRL scores did not signal a particular (lack of) strategy and the behaviour in their process model is diverse. The weak regulators, in contrast, followed the prescribed learning process almost to the letter; there were only few exceptions in the transitions present in their process model. The average regulators, weak regulators, and self-regulators all showed a nonconformity to the intended course structure in the form of a transition from watching the introductory video to answering a recap question correctly. The fact that skipping in all three cases occurred prior to answering questions correctly may suggest that these learners felt like they could already answer the question without further information and indeed were able to do so. The remaining learning process of the weak regulators was highly regulated by the course design. The learners in the self-regulators cluster, on the other hand, showed that they regulated their learning in a manner that suited themselves. This was in line with their self-reported SRL in terms of high scores on metacognitive skills, time management, environmental structuring, and persistence. Browsing and posting on the forum were clearly integrated in their learning process. It appears as if they used the forum as a source of help throughout their entire learning process; sometimes only browsing, but sometimes also posting after browsing. Finally, for learners in the help-seeking cluster, engagement with the course forum was not as integrated as for the self-regulators. This is somewhat counter-intuitive, as the help seekers reported the highest levels of help-seeking on the SRL questionnaire. For the help-seeking cluster however, forum engagement was the only transition after finishing a module and before starting the next module, making it an essential transition in their learning process. Browsing and posting on the forum fitted with their self-reported SRL strategy of looking for help when needed. It was, however, surprising that they did not browse the forum first as that could have been an easier way to find help compared to posting on the forum.

From these findings, we conclude that differences in SRL indeed relate to differences in learner behaviour. Furthermore, differences in scores on specific SRL scales could be related to specific learning processes. We have thereby shown how SRL impacts learner behaviour, providing evidence for the claim posited by Li and Baker (2018) that differences between learners influence course activity. Our results thus support Maldonado-Mahauad et al.'s (2018) suggestion that differences in learning processes are the result of differences in SRL.

Our study mostly resembled the work conducted by Kizilcec et al. (2017), but differed in two ways allowing us to extend their findings. First, we focused on SRL as a construct, taking the correlation between SRL scales into account. Kizilcec et al. (2017) found that learners who reported more SRL, more often revisited course materials (e.g., assessments, lectures) that they already completed. By clustering learners into SRL profiles before exploring the influence of SRL on behaviour, we were able to show that high SRL is related to a much wider range of deviations from the course structure. In other contexts, such variety in learning activities has been found to be associated to increased achievement (Fincham et al., 2018; Hadwin et al., 2007). Learners high in SRL thus seem better able to deal with the autonomy offered in the MOOC. A second difference between the study conducted by Kizilcec et al. (2017) and ours is that we focused on learning as a process instead of a collection of transitions. By analysing learners' online behaviour through process mining we were able to identify the strong influence of the course structure on learner behaviour: The course structure was visible in all process models. Identification of the influence of the course structure would have been much more complicated when analysing individual transitions in learner behaviour, showing the benefit of our approach for analysing learner behaviour.

4.1 | Limitations and suggestions for future research

While the results of the current study increase our knowledge of the influence of SRL on learner behaviour, the study is also subject to a number of limitations. The most prominent limitation of the current study is its sample: participants originated from a single MOOC, sample size was limited, and learners self-selected to fill out the questionnaire and thus to participate in this study. The generalizability of this study is limited due to these sampling issues. However, if participants would have studied in different MOOCs, with different structures, the impact of the course structure on learner behaviour patterns would likely have been obscured if the data had been analysed at once. It would be worthwhile to analyse the impact of SRL on learner behaviour in future studies in diverse contexts and with larger samples. Thereby, it could for instance be determined if weak regulators also exhibit less variety in their behaviour in other MOOCs. It would be especially interesting to study the influence of SRL on learner behaviour in a less structured MOOC, as we found a strong influence of the course structure on learner behaviour in the current study. If our findings can be replicated to different contexts, then SRL can explain (some of) the variability in online learner behaviour.

Furthermore, process mining as a methodology to study learner behaviour in MOOCs has great advantages: It enabled us to analyse the large amounts of event data and to create accompanying visualizations to make the data insightful. We however also identify two main limitations associated with process mining. First, data processing and process mining settings influence the results obtained. Transparent reporting of procedures and the consequences of decisions made during analysis is thus essential. We have therefore reported on our data filtering (i.e., what activities in the trace data were retained) and our process mining settings in the current study. As it is not feasible to compare process models of all individual learners and compare those to their SRL scores, learners had to be clustered. We assumed that learners within each cluster would behave in a similar manner. The variety of transitions represented in a process model is then the result of the variety in learning processes *within* learners. However, the variety of transitions may also be (partly) resulting from a variety in learning processes *between* learners. Additional research studying the extent to which learners with similar SRL also behave similarly is needed. If learners with similar SRL vary in their behaviour in a MOOC, then this variability in behaviour may likely be the consequence of other differences between learners, for instance in motivation or prior knowledge. Zooming in on learners with similar self-reported SRL would help isolate the influence of SRL from the effects of other learner differences on learners' online behaviour.

Second, the analysis of learner behaviour with process mining is limited to the analysis of trace data. Learner behaviour outside of the MOOC environment (e.g., consulting other sources) is not stored and can therefore not be analysed. However, the storage of learner behaviour into trace data in MOOCS is at a very fine granularity (every mouse click) and a long time span (the whole length of the course). MOOC trace data are thereby more complete than any other long term data collection, and as no learning could occur without interacting with the MOOC, all crucial learning activities were included by analysing the trace data.

4.2 | Practical implications

In MOOCs, learners are offered great autonomy over their learning process, making adequate SRL vital for learners (e.g., Azevedo & Aleven, 2013; Beishuizen & Steffens, 2011; Kizilcec & Halawa, 2015; Wang et al., 2013). Learners often struggle to successfully regulate their learning (e.g., Azevedo & Cromley, 2004; Bol & Garner, 2011). Many learners may therefore benefit from SRL support, as SRL support can lead to increased course completion and reduced learner dropout (Hew & Cheung, 2014; Kizilcec & Halawa, 2015; Yeomans & Reich, 2017). Unfortunately, many learners appear unable to successfully monitor their own learning needs and are unable to estimate the benefit they could have from using these tools. Low performing learners are especially unsuccessful at monitoring their need for support, while they are most in need of support and could thus benefit most (Clarebout et al., 2010; Clarebout & Elen, 2006). It has therefore been described that the use of such support tools should be encouraged by embedding the tools within the course environment, instead of providing them optionally (Clarebout et al., 2010; Clarebout & Elen, 2006). The results of the current study indicate that if SRL support would be integrated in the course structure, weak regulators – who likely have the greatest need for SRL support - are expected to come into contact with the support automatically and would thus, hopefully, benefit.

However, implementing SRL support requires balance between stimulating support use and respecting learners' autonomy. While many learners would benefit from SRL support, demanding compliance with an SRL intervention interferes with the open nature of MOOCs. Furthermore, high self-regulating students could be frustrated by mandatory support, leading to negative effects on their motivation and performance (Clarebout et al., 2010; Narciss et al., 2007). While learners in the other three clusters (average regulators, help seekers, self-regulators) deviated more from the intended course structure compared to the weak regulators, the course design was also visible in the process models of these three clusters. We therefore propose that an intervention should be designed in such a way that it may be ignored by learners, to not frustrate high-self regulated learners. Support that is integrated in the course in such a way that is automatically presented to MOOC learners, but that can be skipped when desired, would allow high self-regulating learners to stick to their personally preferred order of learning activities.

Of course, the option to skip support may also be used by learners that would highly benefit from it. Future studies might find it possible to identify learners in need of support based on their behaviour in the online course environment. Potentially, learners in need of SRL support could then be identified during the course and interventions could be implemented only when needed, and tailored to the specific learners' needs. This would provide a solution for the presented conflict between embedding and obligating support for those unable to identify their need for support, and allowing those who are able to regulate their own learning to structure their learning in the way they desire.

Finally, the results of this study also indicate the practical benefit of process mining as a worthwhile addition to the toolkit of course designers. Process mining can provide educational designers with insight on whether learners are following the course structure they designed. For this purpose, process models could also be inspected at a greater level of detail. For instance, by analysing the trace data at the level of individual videos instead of grouping all videos into introduction, content, and summary videos, course designers could identify points in the course where learners often deviate from the intended structure. Course designers could use this information to further develop their online courses.

5 | CONCLUSION

In this study, we investigated the relationship between SRL and learner behaviour in a MOOC. We did so by clustering learners based on their self-reported SRL and comparing the process models of their learning activities. Differences in learner behaviour between the clusters were found, and these differences could be interpreted by using the clusters' SRL scores. Most importantly, weak self-regulated learners had a much more linear approach to studying compared to strong self-regulated learners. The results of this exploratory study show how SRL can influence learner behaviour in a MOOC. We have thereby improved our understanding of the impact of learner heterogeneity on the variety in learner behaviour online. While we acknowledge further research is necessary, our methods and results provide a valuable first step for others to build upon when investigating how SRL impacts learners' online study process.

ACKNOWLEDGMENTS

The authors wish to thank Gwenda Frederiks and Ulrike Wild from Wageningen University and Research for their collaboration. This work was financed via a grant by the Dutch National Initiative for Education Research (NRO)/The Netherlands Organization for Scientific Research (NWO) and the Dutch Ministry of Education, Culture and Science under the grant no. 405-15-705 (SOONER/http://sooner.nu).

CONFLICT OF INTEREST

The authors declare that they have no conflict of interest.

PEER REVIEW

The peer review history for this article is available at https://publons.com/publon/10.1111/jcal.12675.

DATA AVAILABILITY STATEMENT

The data that support the findings of this study are available on request from the corresponding author. The data are not publicly available due to privacy restrictions.

ORCID

Renée S. Jansen https://orcid.org/0000-0002-8385-8322

REFERENCES

Azevedo, R., & Aleven, V. (2013). Metacognition and learning technologies: An overview of current interdisciplinary research. In R. Azevedo & V. Aleven (Eds.), *International handbook of metacognition and learning technologies* (pp. 1–16). New York, NY: Springer Science+Business Media. https://doi.org/10.1007/978-1-4419-5546-3_1

Azevedo, R., & Cromley, J. G. (2004). Does training on self-regulated learning facilitate students' learning with hypermedia? *Journal of Educational Psychology*, *96*, 523–535. https://doi.org/10.1037/0022-0663.96.3.523

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

Barnard-Brak, L., Lan, W., & Paton, V. O. (2010). Profiles in self-regulated learning in the online learning environment. *International Review of Research in Open and Distance Learning*, *11*(1), 61–79.

Beishuizen, J., & Steffens, K. (2011). A conceptual framework for research on self-regulated learning. In R. Carneiro, P. Lefrere, K. Steffens, & J. Underwood (Eds.), *Self-regulated learning in technology enhanced learning environments* (pp. 3–19). Sense Publishers.

Bol, L., & Garner, J. K. (2011). Challenges in supporting self-regulation in distance education environments. *Journal of Computing in Higher Education*, *23*(2–3), 104–123. https://doi.org/10.1007/s12528-011-9046-7

Clarebout, G., & Elen, J. (2006). Tool use in computer-based learning environments: Towards a research framework. *Computers in Human Behavior*, *22*(3), 389–411. https://doi.org/10.1016/j.chb.2004.09.007

Clarebout, G., Horz, H., Schnotz, W., & Elen, J. (2010). The relation between self-regulation and the embedding of support in learning environments. *Educational Technology Research and Development*, *58*, 573–587. https://doi.org/10.1007/s11423-009-9147-4

Davis, D., Triglianos, V., Hauff, C., & Houben, G.-J. (2018). SRLx: A personalized learner interface for MOOCs. In V. Pammer-Schindler, M. Pérez-Sanagustín, H. Drachsler, R. Elferink, & M. Scheffel (Eds.), *Lifelong technology-enhanced learning* (Vol. 11082, pp. 122–135). Switzerland AG: Springer Nature. https://doi.org/10.1007/978-3-319-98572-5_10

Deng, R., Benckendorff, P., & Gannaway, D. (2019). Progress and new directions for teaching and learning in MOOCs. *Computers & Education*, *129*, 48–60. https://doi.org/10.1016/j.compedu.2018.10.019

Dörrenbächer, L., & Perels, F. (2016). Self-regulated learning profiles in college students: Their relationship to achievement, personality, and the effectiveness of an intervention to foster self-regulated learning. *Learning and Individual Differences*, *51*, 229–241. https://doi.org/10.1016/j.lindif.2016.09.015

Fincham, O. E., Gasevic, D. V., Jovanovic, J. M., & Pardo, A. (2018). From study tactics to learning strategies: An analytical method for extracting interpretable representations. *IEEE Transactions on Learning Technologies*, *1–1*, 59–72. https://doi.org/10.1109/TLT.2018.2823317

Garrison, D. R. (2003). Self-directed learning and distance education. In M. G. Moore & W. G. Anderson (Eds.), *Handbook of distance education* (pp. 161–168). Lawrence Erlbaum Associates.

Goda, Y., Yamada, M., Kato, H., Matsuda, T., Saito, Y., & Miyagawa, H. (2015). Procrastination and other learning behavioral types in e-learning and their relationship with learning outcomes. *Learning and Individual Differences*, *37*, 72–80. https://doi.org/10.1016/j.lindif.2014.11.001

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

Hew, K. F., & Cheung, W. S. (2014). Students' and instructors' use of massive open online courses (MOOCs): Motivations and challenges. *Educational Research Review*, *12*, 45–58. https://doi.org/10.1016/j.edurev.2014.05.001

Jansen, R. S., Van Leeuwen, A., Janssen, J., Kester, L., & Kalz, M. (2017). Validation of the self-regulated online learning questionnaire. *Journal of Computing in Higher Education*, *29*, 6–27. https://doi.org/10.1007/s12528-016-9125-x

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

Kizilcec, R. F., & Halawa, S. (2015). *Attrition and achievement gaps in online learning* (pp. 57–66). ACM. https://doi.org/10.1145/2724660.2724680

Kizilcec, R. F., Pérez-Sanagustín, M., & Maldonado, J. J. (2016). *Recommending self-regulated learning strategies does not improve performance in a MOOC* (pp. 101–104). ACM Press.

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

Kizilcec, R. F., Piech, C., & Schneider, E. (2013). *Deconstructing disengagement: Analyzing learner subpopulations in massive open online courses*. Presented at the LAK.

Kovanović, V., Gašević, D., Joksimović, S., Hatala, M., & Adesope, O. (2015). Analytics of communities of inquiry: Effects of learning technology use on cognitive presence in asynchronous online discussions.

The Internet and Higher Education, *27*, 74–89. https://doi.org/10.1016/j.iheduc.2015.06.002

Li, Q., & Baker, R. (2018). The different relationships between engagement and outcomes across participant subgroups in massive open online courses. *Computers & Education*, *127*, 41–65. https://doi.org/10.1016/j.compedu.2018.08.005

Littlejohn, A., Hood, N., Milligan, C., & Mustain, P. (2016). Learning in MOOCs: Motivations and self-regulated learning in MOOCs. *The Internet and Higher Education*, *29*, 40–48. https://doi.org/10.1016/j.iheduc.2015.12.003

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

Milligan, G. W. (1980). An examination of the effect of six types of error perturbation on fifteen clustering algorithms. *Psychometrika*, *45*, 325–342. https://doi.org/10.1007/BF02293907

Mooi, E., & Sarstedt, M. (2010). Cluster analysis. In E. Mooi & M. Sarstedt (Eds.), *A concise guide to market research* (pp. 237–284). Berlin Heidelberg: Springer-Verlag. https://doi.org/10.1007/978-3-642-12541-6_9

Narciss, S., Proske, A., & Koerndle, H. (2007). Promoting self-regulated learning in web-based learning environments. *Computers in Human Behavior*, *23*(3), 1126–1144. https://doi.org/10.1016/j.chb.2006.10.006

Nelson, T. O., & Narens, L. (1990). Metamemory: A Theoretical Framework and New Findings. In *Metamemory: A theoretical framework and new findings. In psychology of learning and motivation* (Vol. 26, pp. 125–173). Academic Press.

Ning, H. K., & Downing, K. (2015). A latent profile analysis of university students' self-regulated learning strategies. *Studies in Higher Education*, *40*(7), 1328–1346. https://doi.org/10.1080/03075079.2014.880832

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

Saint, J., Gašević, D., & Pardo, A. (2018). Detecting learning strategies through process mining. In V. Pammer-Schindler, M. Pérez-Sanagustín, H. Drachsler, R. Elferink, & M. Scheffel (Eds.), *Lifelong technology-enhanced learning* (Vol. 11082, pp. 385–398). Switzerland AG: Springer Nature. https://doi.org/10.1007/978-3-319-98572-5_29

Sitzmann, T., & Ely, K. (2011). A meta-analysis of self-regulated learning in work-related training and educational attainment: What we know and where we need to go. *Psychological Bulletin*, *137*, 421–442. https://doi.org/10.1037/a0022777

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.

Sonnenberg, C., & Bannert, M. (2018). Using process mining to examine the sustainability of instructional support: How stable are the effects of metacognitive prompting on self-regulatory behavior? *Computers in Human Behavior*, *96*, 259–272. https://doi.org/10.1016/j.chb.2018.06.003

Wang, C.-H., Shannon, D. M., & Ross, M. E. (2013). Students' characteristics, self-regulated learning, technology self-efficacy, and course outcomes in online learning. *Distance Education*, *34*, 302–323. https://doi.org/10.1080/01587919.2013.835779

Waschull, S. B. (2001). The online delivery of psychology courses: Attrition, performance, and evaluation. *Teaching of Psychology*, *28*, 143–147.

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

Yeomans, M., & Reich, J. (2017). Planning prompts increase and forecast course completion in massive open online courses. In *Proceedings of the Seventh International Learning Analytics & Knowledge Conference – LAK'17* (pp. 464–473). ACM. https://doi.org/10.1145/3027385.3027416

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

Zimmerman, B. J. (2008). Investigating self-regulation and motivation: Historical background, methodological developments, and future prospects. *American Educational Research Journal*, *45*(1), 166–183. https://doi.org/10.3102/0002831207312909

How to cite this article: Jansen, R. S., van Leeuwen, A., Janssen, J., & Kester, L. (2022). Exploring the link between self-regulated learning and learner behaviour in a massive open online course. *Journal of Computer Assisted Learning*, *38*(4), 993–1004. https://doi.org/10.1111/jcal.12675