Teachers' Learning Profiles in Learning Programming: The Big Picture!

Mohammed Saqr[1]*, Ville Tuominen[2]*, Teemu Valtonen[3], Erkko Sointu[4], Sanna Väisänen[3] and Laura Hirsto[3]

[1] School of Computing, University of Eastern Finland, Joensuu, Finland, [2] Valamis Group Ltd., Joensuu, Finland, [3] School of Applied Educational Science and Teacher Education, University of Eastern Finland, Joensuu, Finland, [4] Special Education, School of Educational Sciences and Psychology, University of Eastern Finland, Joensuu, Finland

Currently there is a need for studying learning strategies within Massive Open Online Courses | (MOOCs), especially in the context of in-service teachers. This study aims to bridge this gap and try to understand how in-service teachers approach and regulate their learning in MOOCs. In particular, it examines the strategies used by the in-service teachers as they study a course on how to teach programming. The study implemented a combination of unsupervised clustering and process mining in a large MOOC (n = 27,538 of which 8,547 completed). The results show similar trends compared to previous studies conducted within MOOCs, indicating that teachers are similar to other groups of students based on their learning strategies. The analysis identified three subgroups (i.e., clusters) with different strategies: (1) *efficient* (n = 3596, 42.1%), (2) *clickers* (n = 1785, 20.9%), and (3) *moderates* (n = 3,166, 37%). The *efficient* students finished the course in a short time, spent more time on each lesson, and moved forward between lessons. The *clickers* took longer to complete the course, repeated the lessons several times, and moved backwards to revise the lessons repeatedly. The *moderates* represented an intermediate approach between the two previous clusters. As such, our findings indicate that a significant fraction within teachers poorly regulate their learning, and therefore, teacher education should emphasize learning strategies and self-regulating learning skills so that teacher can better learn and transfer their skills to students.

Keywords: learning analytics, process mining, in-service teachers, self-regulated learning, learning strategies, clustering, MOOCs, educational data mining

INTRODUCTION

This study targets an in-service teacher training in a Massive Open Online Course | (MOOC) focusing on programming skills. Currently, programming is seen as an important area skill that shows within national curricula and as one of the 21st century skills (Voogt and Roblin, 2012). In developing teachers' skills, the focus has traditionally been on educating teachers through professional development courses (e.g., Hirsto and Löytönen, 2011; Opfer and Pedder, 2011). However, these individual courses have not proved very effective in transforming teachers' practices (Borko, 2004; Korthagen, 2016). Along with traditional professional development

courses, different online courses and webinars provide more flexible ways for in-service training, allowing teachers to adjust their timetables and learning practices to be flexible alongside personal life (e.g., Gordillo et al., 2021). Still, along with flexibility, online learning is seen as taxing and requiring learners to have strong motivation, self-regulation, and efficient learning strategies (Azevedo, 2015). There is an agreement among theoreticians that students with poor self-regulation skills use inappropriate learning strategies, choose poor learning pathways, and often struggle with performing online learning tasks (Zimmerman, 1990; Schunk and Zimmerman, 2012). Whereas students with efficient learning strategies are more likely to complete their courses, remain engaged, and graduate. Especially with MOOCs, the challenge has been indicated by the high dropout rates, as only a minority of students who start the course complete it. Such low completion rate has been attributed to a multitude of factors, e.g., lack of enough time to be able to catch up with MOOC requirements. Other causes include feeling of isolation, lack of motivation, hidden costs, e.g., purchase of books or subscriptions and insufficient background knowledge (Khalil and Ebner, 2014). Such challenges with online courses, and especially with MOOCs, pose questions for learning strategies used by successful students.

Research using learning analytics (LA) has grown rapidly over the past decade. Universities, schools, and companies are actively working to capitalize on the potential that the data can bring (Bergdahl et al., 2020). LA can be defined as "the measurement, collection, analysis and reporting of data about learners and their contexts, for purposes of understanding and optimizing learning and the environments in which it occurs" (Siemens, 2013). The rapid growth in the field of LA has resulted in a vibrant community of researchers with several threads of research that span a wide range of applications (Bergdahl et al., 2020). One such line of research is targeted at "understanding and optimizing learning." In doing so, researchers used analytical methods to investigate students' approaches to learning, to identify learners' self-regulation, collaborative patterns, or to map students' learning processes (López-Pernas et al., 2021, 2022; Malmberg et al., 2022).

A considerable corpus of LA research has been conducted, targeting learners' self-regulation in formal education, where courses are organized by a higher education institution with typical scheduled times, offered to a limited number of students and supported by educators (e.g., Jovanović et al., 2017; López-Pernas et al., 2021). In addition, there is a considerable number of studies focusing on learning strategies used within MOOCs that are open to everyone, regardless of their background skills or training (Joksimović et al., 2018; Kovanović et al., 2019). Nevertheless, MOOCs designed for professional development, and especially for in-service teacher training, have remained largely unstudied. Teachers as professionals of pedagogy and learning pose an interesting research target within a learning context such as MOOCs, which are expected to pose high demands for learners' self-regulations skills. Based on their teacher training, one may assume that they have good starting points for learning within courses demanding strong learning skills, such as MOOCs. Within this study, we focus on this area, namely in-service teachers' learning strategies within a course about teaching coding for students, offered through a professional Massive Open Online Course (pMOOC). This study covers a large course (*n* = 27,538 of which 8,547 completed). The aim is to outline the learning strategies of professional teachers within a pMOOC targeted at learning coding. The aim is to reflect the results from the perspective of self-regulated learning (Jovanović et al., 2017; López-Pernas et al., 2021). The research questions of this study are:

> RQ1: What are the adopted learning strategies that teachers use in pMOOCs?
>
> RQ2: What can process mapping tell us about learners' strategies in pMOOCs?

THEORETICAL BACKGROUND

Self-Regulation and Approaches to Learning

This study builds on the theory of self-regulated learning (Schunk and Zimmerman, 2012) within the context of pMOOCs. According to Boekaerts (1999), SRL is a complex construct positioned at the junction of different overlapping research areas, such as approaches to learning, metacognition, and regulation, along with theories of goal-directed behavior. Within the approaches to learning, two different kinds of categories, deep and surface, have been recognized (Marton and Säljö, 1976; Entwistle et al., 2001). Students presenting deep learning put emphasis on identifying the meanings of the content studied, trying to make connections between new and existing knowledge along with their own experiences. On the other hand, surface-level learners typically rely more on rote memorization, settling for memorizing unrelated pieces of information targeted for the assessment. Following these categories, a strategic approach was identified a little later (Entwistle and Ramsden, 1983).

The SRL is seen as a cyclical process proceeding from forethought to performance, or volitional control to a self-reflection phase (Zimmerman, 1990; Schunk and Zimmerman, 2012). Based on this model, the learning process is outlined as an acknowledged process posing an important position for the learner's conscious decisions and metacognitive thinking. A central aspect of SRL is learners' perception and ability to regulate, monitor, and direct their learning process (Winne, 1995; Jovanović et al., 2021). Previous studies have shown that self-regulation skills correlate with students' course performance and predict students' academic performance, and self-regulation skills were also related to students' help-seeking activity from their peers and teachers, resulting in positive learning results (Peeters et al., 2020; Kashif and Shahid, 2021). The challenge with SRL is that, despite the assumed possibility for acknowledging the learning process, learners are not typically aware of their personal learning practices, as the different learning activities and choices are not based on acknowledged decisions. In order to be a self-regulating learner, one needs to be aware that alternative ways of studying the topic exist. Nevertheless, according to Perry

(1998), learning skills- i.e., skills in using and acknowledging different learning strategies- can be taught to students. The ability to regulate and monitor one's learning process can be seen as an internally regulated process, and along with internal regulation, the learning process can also be seen as an externally regulated process (Boekaerts, 1999; Jovanović et al., 2021). The internal regulation process refers to the learners' own choices and decisions concerning their learning goals and the means used to achieve the goals (Saqr and Lopez-Pernas, 2021). The external regulation refers to learning processes in which the regulations are provided by others, typically teachers or peers, or other available guidelines and materials (Jovanović et al., 2021; Malmberg et al., 2022).

Learning Strategies and Learning Analytics

SRL has typically been studied and measured using measures such as the Motivated Strategies for Learning Questionnaire (MSLQ) (Pintrich et al., 1993) and Learning and Study Strategies Inventory (Weinstein et al., 1987). Instead of using aptitude measures (e.g., MSLQ or LASSI) or rating scales for SRL, i.e., individuals rate themselves using categorical responses such as "most of the time," LA data and its analysis has provided new ways for studying and explaining learners' approaches to learning, and ways to make the learning processes visible. The most commonly used methods are unsupervised clustering, sequence mining, and process mining (Ghazal et al., 2017; Romero and Ventura, 2020). Unsupervised clustering aims to find similarities between learners, learners' actions, and sequences of learners' behaviors. The premise is that the clustering algorithm can aggregate patterns that share similar attributes into interpretable clusters, such as clusters of learners who adopt a similar strategy or clusters of learners who share a similar sequence of actions (Jovanović et al., 2017). Sequence mining (SM) is an analytical approach that deals with sequences of time-ordered categorical data. Temporality and order are at the heart of SM and, therefore, it is always implemented to understand the temporal patterns of learning (López-Pernas et al., 2021; López-Pernas and Saqr, 2021). The third common method is process mining, which maps students' learning actions and transitions between actions, as well as the flow of such actions, using detailed graphical representations (Peeters et al., 2020). It has become common to combine analytical methods to offer a comprehensive perspective of learning behaviors. In such lines of research, sequence mining is typically used to represent and visualize the learning actions; unsupervised clustering is used to group similar patterns of learning sequences; and process mining is used to visualize the clusters of learning sequences (Saqr et al., 2021). The following section will explore previous research that has explored learners' self-regulation using learning analytics methods (Ghazal et al., 2017; López-Pernas et al., 2021).

There have been several studies targeting learning strategies of students in different MOOC courses and in online, flipped, or blended learning courses. Typically, the learning strategies, grounded on perspectives of self-regulated learning, are studied using the activity data from the learning environments (Joksimović et al., 2018; López-Pernas and Saqr, 2021). The data contain information about how different resources and activities such as forums, videos, tutorials, and quizzes are used, and how students are able to study within the schedules provided. Kizilcec et al. (2013) studied engagement of students during three MOOC courses in computer science. Their results provided four different learning strategies, namely: completing, auditing, disengaging, and sampling. Kovanović et al. (2015) were able to infer the presence of six clusters of learner profiles, and such profiles ranged from highly intensive and intensive users who are heavy learners of the platform, to task and content-oriented learners, as well as none-users who barely used the support platform. Jovanović et al. (2017) used a combination of unsupervised clustering and sequence mining to investigate how students regulate their learning in a flipped classroom. The authors found five distinct clusters of learning strategies that represent three levels of engagement: (1) an intensive cluster, (2) a strategic and highly strategic cluster, and (3) a selective and highly selective cluster. Taken together, the results show similar features, at one end are students who take advantage of the various possibilities, materials, and activities provided to support their learning, and at the other end are students who enroll on the course but do not fully use the possibilities provided to support learning, or do not proceed further on the course. In addition, the roles of different assessment activities and the use of discussions and videos cause differences among students' learning strategies. We assume that these results reflect the approaches to learning, namely deep and surface levels (Entwistle and Ramsden, 1983; Marton and Säljö, 1997), and abilities for self-regulation and internal and external regulation (Boekaerts, 1999). Within this study, we proceed with this process by targeting professional educators, in-service teachers. The aim is to provide insights into their learning strategies reflected *via* SRL (**Figure 1**).

MATERIALS AND METHODS

Context

The course is programming in societies for in-service teachers. The main page guides the learners and introduces them to the topic and the studies. The studies are built and presented as a linear path of eight content chapters (referred to as *lessons*). Lessons are rich in multimedia and links to other pages, and two of them include quizzes. The order of progress is free. For quizzes, the pass score is 70%. Completion of all lessons grants a certificate.

Lesson Contents

1. Programming in societies: the lesson introduces the concept of programming and coding. What those have meant in the past and what they mean now. Students need to complete the self-evaluation form and multichoice questionnaire. Students need to answer more than 70% of the questions correctly.
2. Try programming: the lesson introduces simple programming techniques and online tools. Students

FIGURE 1 | Course main interface, which shows the sequence of the lessons.
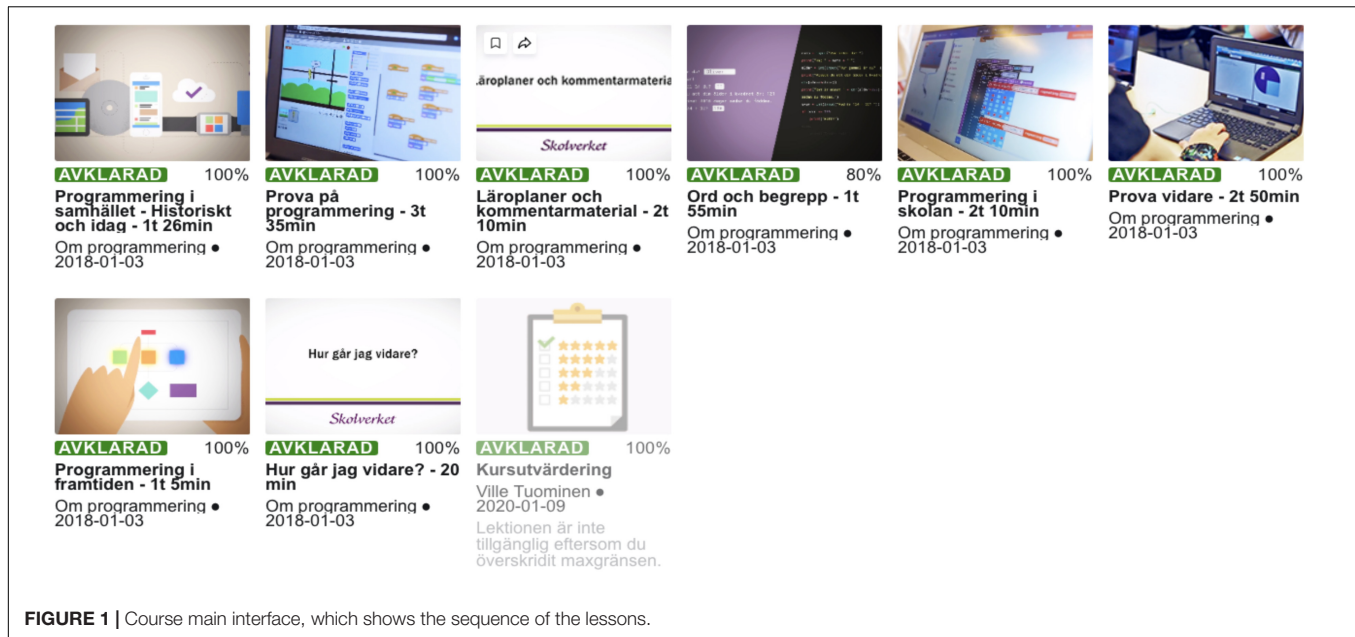
are guided to try those tools and comment on the public channel about the experience.

3. How to learn programming: the lesson introduces the digitalization and programming curricula at different school levels. Students need to get to know the material and answer questions in the forum.

4. Terminology: the lesson introduces the key concepts and terminology of coding and programming. Students need to pass the multichoice quiz with a 70% pass score.

5. Programming in schools: the lesson introduces students to programming practicalities at different school levels. Students need to get to know the material and reflect in the forum.

6. Try more: the lesson introduces more programming techniques and tools. Students are guided to apply those techniques and tools during the lesson. At the end of the lesson, students reflect in the forum.

7. Future development of programming: the lesson introduces the question "what does the future look like, how does programming affect us?" Students need to get to know the material and reflect in the forum.

8. How do I move on?: the lesson introduces other resources to which students can move on after this course. Students answer a self-evaluation, and the course feedback questions.

Data Collection

Data was collected from the Learning Management System (LMS) for each student. The LMS contained data for 27,538 students. Only students who completed the course were included ($n = 8,547$ completed). The log data contained 4,462,458 data records, which was filtered to 2,265,655 after excluding the students who did not complete the course. All students completed the eight lessons of the course. The following variables were calculated for each

student, to capture their online behavior: (1) N lessons, (2) N successful lessons, (3) video, (4) total view, (5) course evaluation, (6) span, (7) duration, and (8) average lesson time.

[1] *The number of lessons* is the number of times a student took a lesson, including repetitions. This indicator is operationalized as a student's interest in the content, looking up information, or revising a section of the lesson. The indicator was included because we noticed a vast difference between the number of lessons takes and successful completion. [2] *The number of successful lessons* indicates the raw total lessons a student completed successfully. Students are likely to complete a lesson successfully when they are more engaged with the content or interested in finishing the course. Please note that all students successfully completed the eight unique lessons of the course. [3] *Video* is the total time a student spent watching the course videos. [4] *Total views* represents the sum of clicks a student made in the LMS. [5] *Course evaluation* represents the number of clicks a student made in the course evaluation module. Note that most of the responses were free text and therefore needed qualitative analysis. [6] *Span* is the difference between the first and last login to the course. [7] *Duration* is the total duration of time the student spent taking the course, meaning the time from start to finish. [8] *Average lesson time* is the average duration a student spent studying a lesson.

Data Analysis

These eight indicators from the LMS were calculated for each student and were used for the clustering of the data. The data were prepared for clustering as follows: outliers were Winsorized (observations beyond the 95th percentile), all variables were standardized ($M = 0$, $SD = 1$), centered (M divided by SD), and checked for collinearity. The K-means clustering method was used, as it has been proven effective in clustering educational data (Joksimović et al., 2018). The

optimum number of clusters was chosen based on the NbClust method of the NbClust R package (Malika et al., 2014). NbClust offers 30 metrics, and each metric suggests an optimum number. Three clusters was the number suggested by most metrics (13 [43.3%]). To test the quality of clustering, the silhouette method was used to test how well separated the clusters are. The silhouette value was 0.43. Furthermore, the three clusters were compared using the Kruskal–Wallis non-parametric one-way analysis of variance (ANOVA) (Ostertagová et al., 2014). To test the magnitude of the ANOVA results and the quality of separation of clusters, we computed the effect size (epsilon-squared [ES]). The interpretation of epsilon-squared followed (Rea and Parker, 2014) with ES values of 0.01 or less as very small, $0.01 \leq ES < 0.06$ – small, $0.16 \leq ES < 0.14$ – medium, $ES \geq 0.14$ – large. *Post hoc* pairwise comparisons were also performed with Dunn's test, using Holm's correction for multiple testing.

The final cluster data were further analyzed with process mining to visualize the students' learning process for the three clusters. The process map was generated using the R library Bupar, which generates visual maps in which the learning activities represent the nodes, and the edges are the frequency of transitions between such activities, meaning how frequently a student moves from, for example, Lesson 1 to Lesson 2 (Janssenswillen et al., 2019). The plotted process was constructed using the timestamp of each click in every lesson; the activity ID was the lesson name; and the frequencies of edges and nodes were based on the relative case frequency, or in other words, the proportion of students who performed an action (López-Pernas et al., 2021).

RESULTS

Using LMS indicators to cluster students into possible homogenous groups with shared patterns of online learning behaviors resulted in three distinct clusters of profiles (**Table 1** and **Figure 2**). We refer to them here as clickers, moderates, and efficient, based on their activities profile.

Clickers ($n = 1,785$, 20.9%): Clickers were a subgroup of learners who clicked on learning resources (e.g., videos) more than any other group, and repeated more lessons over a long time span (the difference between their starting date and finishing the course). In each lesson, on average, they spent the shortest time per lesson among all learners. This could be explained by the fact that they accessed lessons to look up information or revise a concept or to answer a question. The clickers were the group that submitted the fewest course evaluations.

Efficient ($n = 3,596$, 42.1%): This subgroup of students was the opposite of the clickers in that they took the course over a short period (span), they did fewer lesson repetitions, they spent more time on each lesson, and the group had the most course completions.

Moderate ($n = 3,166$, 37%): Learners in this subgroup were moderate in their approach, based on the values from their usage of the learning management system. They spent less time on each lesson than average, and they finished the lessons within a short time span. Their activities lie midway between the clickers and efficient subgroups. This group also posted the most course evaluations.

To test the separation of the clusters, KW ANOVA was performed to compare the standardized values of the indicators. The results show that the mean values of each indicator differed significantly across the three clusters, with a large effect size in all indicators except for course evaluation, which had a moderate effect size (**Table 2**). All pairwise comparisons of clusters for each indicator were statistically significant except for course evaluation. Such findings strongly support the conclusions that the clusters were well separated and had distinct characteristics.

Process Mining

In this section, we use process maps to understand the strategies adopted by several learners and how they navigate their learning process.

Clickers

The process map of clickers (**Figure 3**) was repetitive, moving forward and backward more than any other group. About half (51%) of the clickers returned to *Lesson 1* from *Lesson 2*. Later, about 20% returned to *Lesson 2* from *Lesson 3,* and 16% to *Lesson 1*. The same approach continued with *Lesson 4,* from which about 17% of the students returned to *Lesson 3*, 12% to *Lesson 2*, and 17% to *Lesson 1*. The roaming behavior continued throughout the course.

Efficient Users

The process maps (**Figure 4**) of efficient users show that they spent less time on the lessons (did not repeat them many times). They moved forward from one lesson to the next in a regular way, and only a few times did they return to review a lesson or look at the answers. This happened mostly in the early lessons, so that they returned from *Lesson 2* to *Lesson 1* in 13% of the sessions. It is also notable how frequently they returned from *Lesson 8*, "How do I move on," from which they returned to *Lesson 4*, "Terminology" (7.4%), and *Lesson 1*, "Programming in societies" (6.5%), which included a self-evaluation form and a multiple-choice questionnaire.

Moderate Users

The moderates (**Figure 5**) moved forward from lesson to lesson, but compared to the efficient users, their rates of returning were higher and more diverse. For example, 30% returned to *Lesson 1*, and similarly, 10% from *Lesson 3* to *Lesson 2*, and 7.64% from *Lesson 2* to *Lesson 1*.

DISCUSSION

This paper aims at providing an understanding of how in-service teachers approach and regulate their learning during a professional MOOC. In particular, it examines the strategies used by these learners when they study programming. The study implemented a combination of unsupervised clustering and

TABLE 1 | Comparison of the standardized values of each indicator across the three clusters.

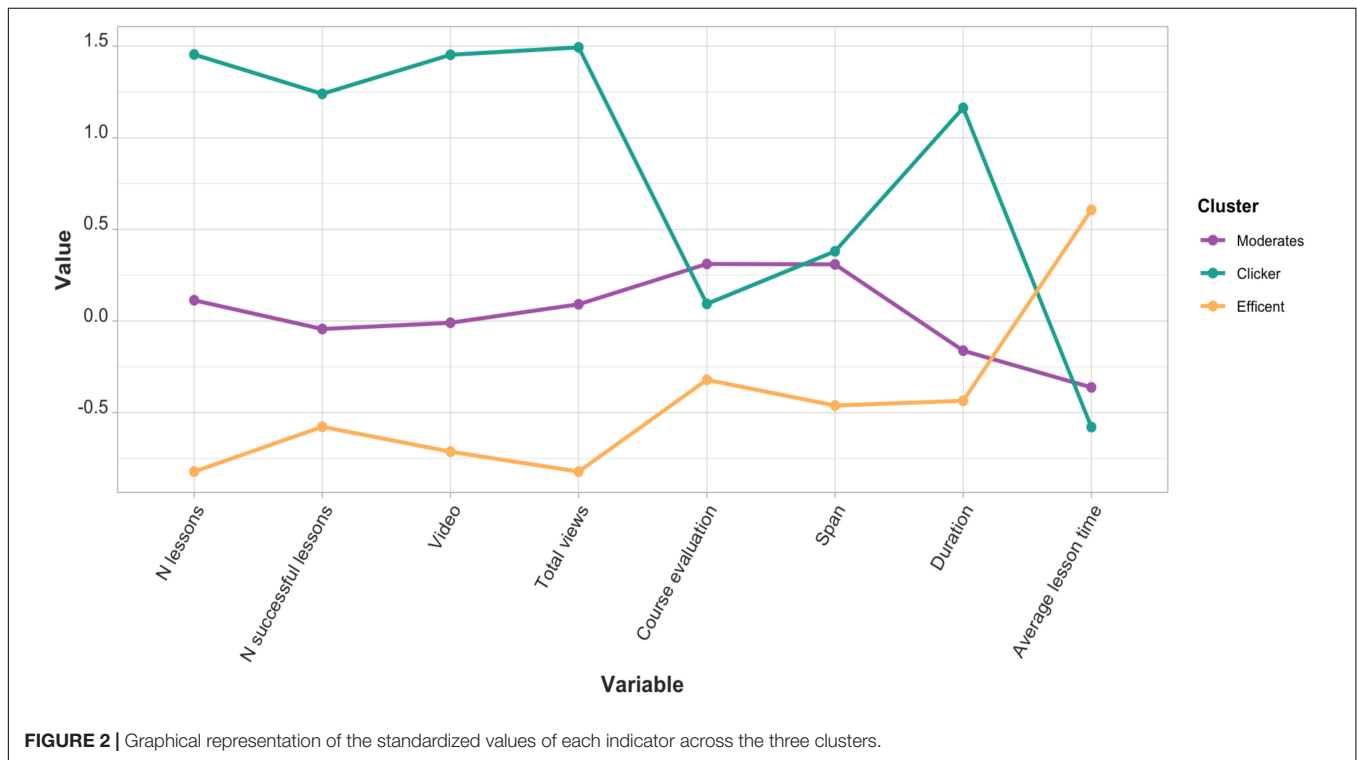| Cluster		N lessons	N successful lessons	Video	Total views	Course evaluation	Span	Duration	Average lesson time
Clicker	Mean	1.45	1.24	1.45	1.49	0.09	0.38	1.16	−0.58
	SD	0.61	1.17	0.71	0.55	1.06	1.11	0.85	0.65
Moderate	Mean	0.11	−0.04	−0.01	0.09	0.31	0.31	−0.16	−0.36
	SD	0.60	0.71	0.66	0.57	1.16	1.11	0.80	0.72
Efficient	Mean	−0.82	−0.58	−0.71	−0.82	−0.32	−0.46	−0.44	0.61
	SD	0.37	0.40	0.44	0.39	0.67	0.57	0.75	1.03

FIGURE 2 | Graphical representation of the standardized values of each indicator across the three clusters.

TABLE 2 | KW ANOVA comparison of the three clusters, showing a well separated clusters with significant differences among each pair of clusters with predominantly large effect size.
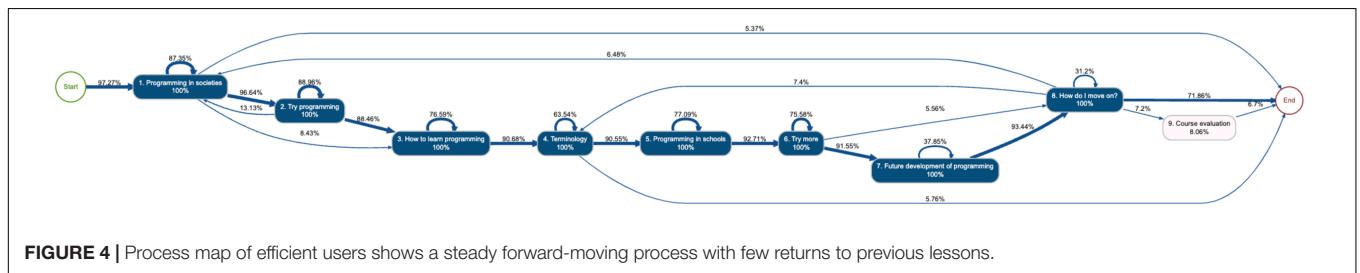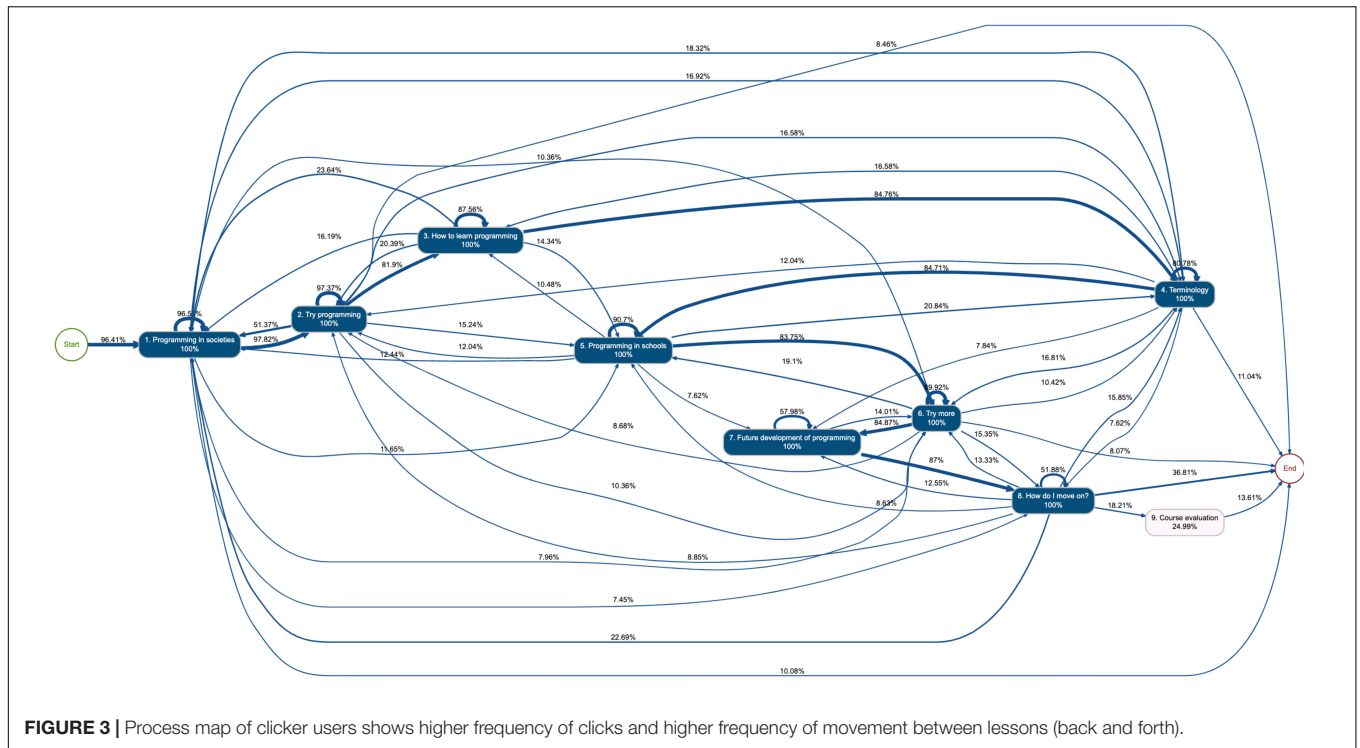
	χ^2	p	ε^2
N lessons*	6,071	<0.001	0.7104
N successful lessons *	3,634	<0.001	0.4252
Video*	5,006	<0.001	0.5857
Total views*	6,158	<0.001	0.7205
Course evaluation	693	<0.001	0.0811
Span*	1,489	<0.001	0.1743
Duration*	2,926	<0.001	0.3424
Average lesson time*	2,446	<0.001	0.2863

Indicates that pairwise comparisons between the pairs of clusters were statistically significant.

process mining. The results of this papers indicate that among in-service teachers with several years of teacher training, there were gaps among teachers' learning strategies.

Our results show three subgroups who ranged from teachers with efficient strategies to the teachers whose actions seemed fumbling, based on trials and errors and another intermediate group. The first subgroup, with an efficient strategy, was the *efficient* group, who completed the course in a short period of time, spent more time on their lessons, and moved forward from one lesson to the next efficiently without going back and forth to revise or look up information. In other words, these *efficient* students spent a focused time learning the content, watching the

FIGURE 3 | Process map of clicker users shows higher frequency of clicks and higher frequency of movement between lessons (back and forth).

FIGURE 4 | Process map of efficient users shows a steady forward-moving process with few returns to previous lessons.
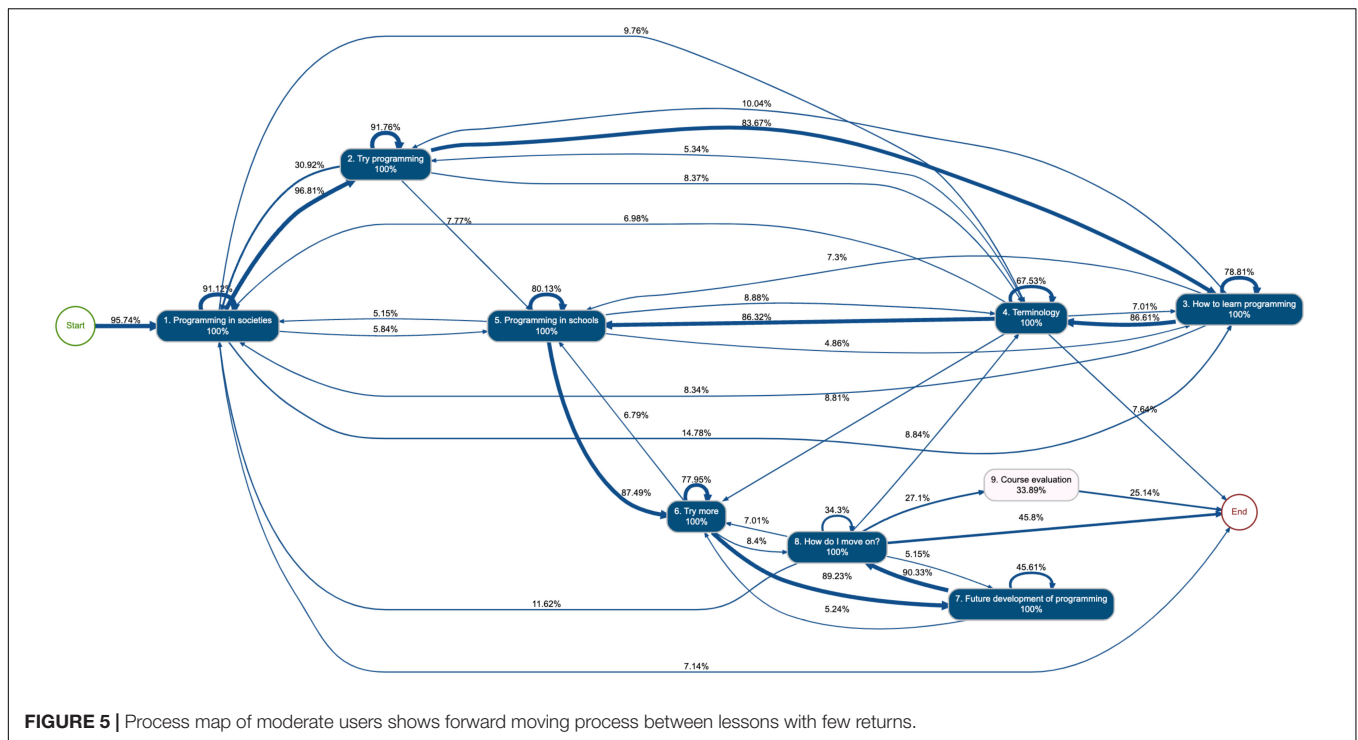
videos, and building on what they had learned from previous lessons. Such a group of learners is similar to the group described by López-Pernas et al. (2021) as strategists. The strategists in the López-Pernas et al. (2021) study were higher education students learning programming, and they made focused efforts with the least number of mistakes and in a short time. In other contexts, such groups of students are similar to *get-it-done* or the strategic group of students described by Kovanović et al. (2015, 2019) and Jovanović et al. (2017), in that they spent less time and were able to pass their courses. As such, the *efficient* students have reasonable strategies and may be the ones who need least support or help during their professional training or life-long learning.

In contrast to the *efficient* group, and at the other end of the spectrum, was the second subgroup, the *clickers*. They clicked intensively and spent large amounts of time on the course and over a long duration. This group of students implemented a technique of trial, error, and revision, in which they seemed to try to guess the answers repeatedly. Most times, this fails, and less often, it works. Thus, they had to go back and revise. The same behavior has been captured by López-Pernas et al. (2021) in a similar context, and they referred to these students as *determined*.

That is, they are trying intensively to get their assignments done. This behavior is different from most of the "*intense*" or "*active*" or "*engaged*" students in other research on strategies and profiles (e.g., Kovanović et al., 2015; Jovanović et al., 2017), in which the intense categories were essentially highly engaged students who scored higher and completed the course on time. In other words, the similarities between our findings and those of López-Pernas et al. (2021) (same programming context), as well as the dissimilarities with other research (different context), emphasize the importance of context. Therefore, a different approach to supporting these students is needed. In fact, helping these students to acquire better regulating strategies could help save their time and institutional resources, and may help them to transfer such strategies to their students.

The third subgroup among our students was the *moderates*, who spent more time than the *efficient* students, but managed to get their lessons completed in a reasonable time. The approach of these students lies in an intermediate zone, meaning that they are less focused and take a longer time to study, and most finish the course. This intermediate category has been described in most previous studies (e.g., Kovanović et al., 2015; Jovanović et al.,

FIGURE 5 | Process map of moderate users shows forward moving process between lessons with few returns.

2017). Since our study included only students who completed the course, which was completely online, we did not have a category for disengaged or isolated students, as in previous research in higher education (e.g., López-Pernas et al., 2021).

The general conclusions from this study revolve around the contextual peculiarities of teaching and learning programming, which necessitates a different approach to course design, instruction, and support. Several red flags were obvious in this study for the clicker group. For example, they went back to the lessons and revised the lessons repeatedly. We assume the explanation for this is that programming environments require support that was not provided by the LMS, such as code snippets, explanations of syntax, or more examples of code. In fact, most LMSs do not offer such resources, and programming learners always resort to websites like StackExchange.com (López-Pernas et al., 2021).

This study provides an important insight into the teachers' learning processes. It seems that teachers, as experts in pedagogy and learning do not show up as a homogenous group with strong skills in regulating their learning, as one might have expected. Students within the efficient cluster proceeded rather directly from the beginning to the end, and they were able to complete the course quite efficiently. We assume the result reflects good readiness for internal self-regulation, setting personal goals and defining strategies to achieve them, and the ability to effectively use the contents and instructions provided within the learning environment. At the other end, the clickers seem to proceed by trying and repeating, reflecting a process that is neither proactive nor built on internal regulation. Instead, we assume that the sequence is based rather on external regulation, searching for information on how

to proceed. The distinction between the efficient group and the clickers may also reflect the different learning strategies and motives. Jumping between lessons and active testing may reflect a surface learning strategy (Entwistle and Ramsden, 1983; Marton and Säljö, 1997; Parpala et al., 2010) and thus the need only to obtain the approved course performance, not to learn the coding.

According to Järvelä and Niemivirta (1999) research on self-regulated learning has typically focused on identifying students with good skills for self-regulation and also on ways to develop students' self-regulation skills. Studies, such as this conducted with analytics, typically focuses on identifying different learning strategies, reflecting self-regulation skills. Within the future the analytics, methods for making the learning strategies explicit, need to be used for longitudinal approaches in the context of developing student's self-regulation skills (e.g., Saqr and López-Pernas, 2021). Further research could also explore the opportunities of idiographic ($n = 1$) learning strategies (e.g., Malmberg et al., 2022), or use replication to verify current findings (e.g., Saqr and López-Pernas, 2022). We assume that analytics would provide tools for following the changes and possible development within strategies and self- regulation skills. This would be important especially within the context of teacher training, aiming at educating new teachers with a good capability to act as a learner with good self-regulation skills and well working learning strategies and to capability to provide these abilities for their students.

These results provide valuable information for further developing MOOCs and other self-study approaches and environments, to support teachers' professional development and

the need for support activities. Perhaps understanding learners' strategies more fully would help us in developing more efficient learning modules that would work powerfully as part of the toolbox to facilitate better professional development for teachers. Despite the limitations, we find these results to be important in developing this research area further using the possibilities of the analytics.

LIMITATIONS AND FUTURE RESEARCH

Nevertheless, lots of possibilities still exist to explain the results. This data did not contain information about the teachers' prior skills in coding and their need for this course. There may have been participants with strong previous coding skills who need only a diploma to get their skills acknowledged. These participants may show as participants with high self-regulation skills, proceeding through the lessons from first to last without returns. This is an area that needs further research, to open the starting levels of the participants and study their relations compared to the cluster. In further research, it would be important to combine more traditional perspectives on in-service teachers' self-regulated learning strategies and approaches to learning with traditional questionnaires, as part of extended or dispositional learning analytics. This would help us to understand learners' behavior in these LMSs more thoroughly in in-service teacher learning. There is also a need for research that can combine the results, and certain phases in the process map, with certain pedagogical designs within the LMS. This would provide us with accurate and targeted information about how certain pedagogical designs work with students with different learning strategies and abilities to regulate their learning. While we used a large dataset, the generalizability of this study may be limited to online programming learning for similar group of students.

CONCLUSION

The area of professional development and, in-service teachers in particular, is unexplored in the literature. This paper addresses such gap and studies the learning strategies of

in-service teachers in a programming course. The results indicate that although teachers have several years of teacher training, a significant fraction of them poorly regulated their online learning. Cluster analysis revealed a heterogenous groups of learners with distinct profiles. An efficient group ($n = 3,596$, 42.1%) who spent more time learning, progressed steadily, and completed the course in short time. A clicker group ($n = 1,785$, 20.9%) who intensely used trial, error and revise strategy and took longer time to finish the course. A third group was ($n = 3,166$, 37%) who are intermediate between the two groups. While a "disengaged" group is commonly reported in the literature, our study found a different "clicker" group who are course-completion or credit oriented.

DATA AVAILABILITY STATEMENT

The datasets presented in this article are not readily available because the data is subject to approval by the respective authorities since it contains private information. Requests to access the datasets should be directed to MS mohammed.saqr@uef.fi.

AUTHOR CONTRIBUTIONS

VT curated the data and prepared the data for analysis. MS, VT, and LH contributed to the analysis. MS contributed to the methods. MS, VT, TV, ES, and LH conceptualized the manuscript. MS, VT, TV, ES, LH, and SV contributed to the writing, revision, and preparing of the manuscript.

FUNDING

REFERENCES

Azevedo, R. (2015). Defining and measuring engagement and learning in science: conceptual, theoretical, methodological, and analytical issues. *Educ. Psychol.* 50, 84–94. doi: 10.1080/00461520.2015.1004069

Bergdahl, N., Nouri, J., Karunaratne, T., Afzaal, M., and Saqr, M. (2020). Learning analytics for blended learning-a systematic review of theory, methodology, and ethical considerations. *Int. J. Learn. Anal. Artif. Intell. Educ.* 2, 46–79. doi: 10.3991/ijai.v2i2.17887

Boekaerts, M. (1999). Self-regulated learning: where we are today. *Int. J. Educ. Res.* 31, 445–457. doi: 10.1016/s0883-0355(99)00014-2

Borko, H. (2004). Professional development and teacher learning: mapping the terrain. *Educ. Res.* 33, 3–15. doi: 10.3102/0013189x033008003

Entwistle, N., McCune, V., and Walker, P. (2001). "Conceptions, styles, and approaches within higher education: analytical abstractions and everyday experience," in *Perspectives on Thinking, Learning, and Cognitive Styles,* eds

R. J. Sternberg and L.-F. Zhang (Mahwah, NJ: Lawrence Erlbaum Associates Publishers), 103–136.

Entwistle, N. J., and Ramsden, P. (1983). *Understanding Student Learning.* London: Croom Helm.

Ghazal, M. A., Ibrahim, O., and Salama, M. A. (2017). "Educational process mining: a systematic literature review," in *Proceedings of the 2017 European Conference on Electrical Engineering and Computer Science (EECS),* Bern, 198–203. doi: 10.1109/EECS.2017.45

Gordillo, A., Barra, E., Garaizar, P., and López-Pernas, S. (2021). "Use of a simulated social network as an educational tool to enhance teacher digital competence," in *IEEE Revista Iberoamericana de Tecnologias del Aprendizaje,* Vol. 16. (Piscataway, NJ: IEEE), 107–114. doi: 10.1109/RITA.2021.3052686

Hirsto, L., and Löytönen, T. (2011). Kehittämisen kolmas tila? Yliopisto-opetus kehittämisen kohteena. *Aikuiskasvatus* 31, 255–266. doi: 10.33336/aik.93953

Janssenswillen, G., Depaire, B., Swennen, M., Jans, M., and Vanhoof, K. (2019). bupaR: enabling reproducible business process analysis. *Knowl. Based Syst.* 163, 927–930. doi: 10.1016/j.knosys.2018.10.018

Järvelä, S., and Niemivirta, M. (1999). The changes in learning theory and the topicality of the recent research on motivation. *Learn. Instr.* 9, 57–65.

Joksimović, S., Poquet, O., Kovanović, V., Dowell, N., Mills, C., Gašević, D., et al. (2018). How do we model learning at scale? A systematic review of research on MOOCs. *Rev. Educ. Res.* 88, 43–86. doi: 10.3102/0034654317740335

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

Jovanović, J., Saqr, M., Joksimović, S., and Gašević, D. (2021). Students matter the most in learning analytics: the effects of internal and instructional conditions in predicting academic success. *Comput. Educ.* 172:104251. doi: 10.1016/j.compedu.2021.104251

Kashif, M. F., and Shahid, R. (2021). Students' self-regulation in online learning and its effect on their academic achievement. *Glob. Educ. Stud. Rev.* VI, 11–20. doi: 10.31703/gesr.2021(vi-iii).02

Khalil, H., and Ebner, M. (2014). "MOOCs completion rates and possible methods to improve retention-a literature review," in *Proceedings of the World Conference on Educational Multimedia, Hypermedia and Telecommunications, 23 Jun 2014 – 26 Jun 2014,* Tampere, 1236–1244.

Kizilcec, R. F., Piech, C., and Schneider, E. (2013). "Deconstructing disengagement: analyzing learner subpopulations in massive open online courses," in *Proceedings of the Third International Conference on Learning Analytics and Knowledge - LAK '13. the Third International Conference,* (Leuven). doi: 10.1145/2460296.2460330

Korthagen, F. (2016). Inconvenient truths about teacher learning: towards professional development 3.0. *Teach. Teach.* 23, 1–19. doi: 10.1080/13540602.2016.1211523

Kovanović, V., Gašević, D., Joksimović, S., Hatala, M., and Adesope, O. (2015). Analytics of communities of inquiry: effects of learning technology use on cognitive presence in asynchronous online discussions. *Internet High. Educ.* 27, 74–89. doi: 10.1016/j.iheduc.2015.06.002

Kovanović, V., Joksimović, S., Poquet, O., Hennis, T., de Vries, P., Hatala, M., et al. (2019). Examining communities of inquiry in massive open online courses: the role of study strategies. *Internet High. Educ.* 40, 20–43. doi: 10.1016/j.iheduc.2018.09.001

López-Pernas, S., and Saqr, M. (2021). Bringing synchrony and clarity to complex multi-channel data: a learning analytics study in programming education. *IEEE Access* 9, 166531–166541. doi: 10.1109/ACCESS.2021.3134844

López-Pernas, S., Saqr, M., Gordillo, A., and Barra, E. (2022). A learning analytics perspective on educational escape rooms. *Interact. Learn. Environ.* 1–17. doi: 10.1080/10494820.2022.2041045

López-Pernas, S., Saqr, M., and Viberg, O. (2021). Putting it all together: combining learning analytics methods and data sources to understand students' approaches to learning programming. *Sustainability* 13:4825. doi: 10.20944/preprints202104.0404.v1

Malika, C., Ghazzali, N., Boiteau, V., and Niknafs, A. (2014). NbClust: an R package for determining the relevant number of clusters in a data set. *J. Stat. Softw.* 61, 1–36.

Malmberg, J., Saqr, M., Järvenoja, H., and Järvelä, S. (2022). How the monitoring events of individual students are associated with phases of regulation. *J. Learn. Anal.* 9, 77–92. doi: 10.18608/jla.2022.7429

Marton, F., and Säljö, R. (1976). On qualitative differences in learning: I-outcome and process. *Br. J. Educ. Psychol.* 46, 4–11. doi: 10.1111/j.2044-8279.1976.tb02980.x

Marton, F., and Saljo, R. (1997). "Approaches to learning," in *The Experience of Learning. Implications for Teaching and Studying in Higher Education,* 2nd Edn. eds F. Marton, D. Hounsell, and N. J. Entwistle (Edinburgh: Scottish Academic Press), 39–58.

Opfer, V. D., and Pedder, D. (2011). Conceptualizing teacher professional learning. *Rev. Educ. Res.* 81, 376–407. doi: 10.3102/0034654311413609

Ostertagová, E., Ostertag, O., and Kováè, J. (2014). Methodology and application of the Kruskal-Wallis test. *Appl. Mech. Mater.* 611, 115–120. doi: 10.4028/www.scientific.net/amm.611.115

Parpala, A., Lindblom-Ylänne, S., Komulainen, E., Litmanen, T., and Hirsto, L. (2010). Students' approaches to learning and their experiences of the teaching-learning environment in different disciplines. *Br. J. Educ. Psychol.* 80, 269–282. doi: 10.1348/000709909X476946

Peeters, W., Saqr, M., and Viberg, O. (2020). "Applying learning analytics to map students' self-regulated learning tactics in an academic writing course," in *Proceedings of the 28th International Conference on Computers in Education,* Vol. 1, eds H.-J. So, M. M. Rodrigo, J. Mason, and A. Mitrovic 245–254.

Perry, N. E. (1998). Young children's self-regulated learning and contexts that support it. *J. Educ. Psychol.* 90, 715–729. doi: 10.1037/0022-0663.90.4.715

Pintrich, P. R., Smith, D. A. F., Garcia, T., and Mckeachie, W. J. (1993). Reliability and predictive validity of the motivated strategies for learning questionnaire (mslq). *Educ. Psychol. Meas.* 53, 801–813. doi: 10.1177/0013164493053003024

Rea, L. M., and Parker, R. A. (2014). *Designing and Conducting Survey Research: A Comprehensive Guide.* Hoboken, NJ: John Wiley & Sons.

Romero, C., and Ventura, S. (2020). Educational data mining and learning analytics: an updated survey. *WIREs Data Min. Knowl. Discov.* 10:e1355. doi: 10.1002/widm.1355

Saqr, M., and López-Pernas, S. (2021). The longitudinal trajectories of online engagement over a full program. *Comput. Educ.* 175:104325. doi: 10.1016/j.compedu.2021.104325

Saqr, M., and Lopez-Pernas, S. (2021). "Idiographic learning analytics: a single student (N=1) approach using psychological networks," in *Companion Proceedings 11th International Conference on Learning Analytics & Knowledge (LAK21),* 397–404.

Saqr, M., and López-Pernas, S. (2022). The curious case of centrality measures: a large-scale empirical investigation. *J. Learn. Anal.* 9, 13–31. doi: 10.18608/jla.2022.7415

Saqr, M., Peeters, W., and Viberg, O. (2021). The relational, co-temporal, contemporaneous, and longitudinal dynamics of self-regulation for academic writing. *Res. Pract. Technol. Enhanc. Learn.* 16:29. doi: 10.1186/s41039-021-00175-7

Schunk, D. H., and Zimmerman, B. J. (2012). *Self-Regulation and Learning. In Handbook of Psychology,* 2nd Edn. Hoboken, NJ: John Wiley & Sons, Inc. doi: 10.1002/9781118133880.hop207003

Siemens, G. (2013). Learning analytics: the emergence of a discipline. *Am. Behav. Sci.* 57, 1380–1400. doi: 10.1177/0002764213498851

Voogt, J., and Roblin, N. P. (2012). A comparative analysis of international frameworks for 21stcentury competences: implications for national curriculum policies. *J. Curric. Stud.* 44, 299–321. doi: 10.1080/00220272.2012.668938

Weinstein, C. E., Schulte, A. C., and Palmer, D. R. (1987). *LASSI: Learning and Study Strategies Inventory.* Clearwater, FL: H & h publishing Company.

Winne, P. H. (1995). Self-regulation is ubiquitous but its forms vary with knowledge. *Educ. Psychol.* 30, 223–228. doi: 10.1207/s15326985ep3004_9

Zimmerman, B. J. (1990). Self-regulated learning and academic achievement: an overview. *Educ. Psychol.* 25, 3–17. doi: 10.1207/s15326985ep2501_2