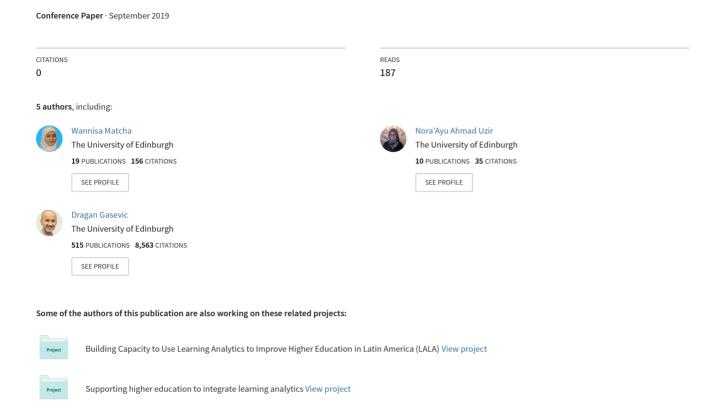
# Detection of Learning Strategies: A Comparison of Process, Sequence and Network Analytic Approaches



# Detection of Learning Strategies: A Comparison of Process, Sequence and Network Analytic Approaches

Wannisa Matcha  $^{1[0000-0003-1120-3324]}$ , Dragan Gašević  $^{1,\,2[0000-0001-9265-1908]}$ , Nora'ayu Ahmad Uzir $^{1,3[0000-0002-9589-2800]}$ , Jelena Jovanović  $^{4[0000-0002-1904-0446]}$ , Abelardo Pardo  $^{5[0000-0002-6857-0582]}$ , Jorge Maldonado-Mahauad  $^{6,7[0000-0003-1953-390X]}$  and Mar Pérez-Sanagustín  $^{6,8[0000-0001-9854-9963]}$ 

<sup>1</sup> University of Edinburgh, Edinburgh, EH8 9AB {w.matcha, n.uzir}@ed.ac.uk
<sup>2</sup> Monash University, Clayton VIC 3800 Australia dragan.gasevic@monash.edu
<sup>3</sup> Universiti Teknologi MARA, 40150, Shah Alam, Malaysia
<sup>4</sup> University of Belgrade, Belgrade 11000 Serbia jelena.jovanovic@fon.bg.ac.rs
<sup>5</sup> University of South Australia, South Australia 5000 Australia
abelardo.pardo@unisa.edu.au
<sup>6</sup> Pontificia Universidad católica de Chile, Santiago Chile
<sup>7</sup> Universidad de Cuenca, Cuenca Ecuador jjmaldonado@uc.cl

<sup>8</sup>Université Paul Sabatier Toulouse III, Institute de Recherche en Informatique de Toulouse (IRIT), F-31062 Toulouse CEDEX 9 mar.perez@ing.puc.cl

Abstract. Research in learning analytics proposed different computational techniques to detect learning tactics and strategies adopted by learners in digital environments through the analysis of students' trace data. While many promising insights have been produced, there has been much less understanding about how and to what extent different data analytic approaches influence results. This paper presents a comparison of three analytic approaches including process, sequence, and network approaches for detection of learning tactics and strategies. The analysis was performed on a dataset collected in a massive open online course on software programming. All three approaches produced four tactics and three strategy groups. The tactics detected by using the sequence analysis approach differed from those identified by the other two methods. The process and network analytic approaches had more than 66 percent of similarity in the detected tactics.

Learning strategies detected by the three approaches proved to be highly similar.

Keywords: learning strategy, learning analytics, data analytics

#### 1 Introduction

The objective of massive open online courses (MOOCs) is to offer learning opportunities to a wide range of learners. However, MOOCs have been associated with high dropout and failure rates [1, 2]. Research identified several factors associated with such course outcomes including motivation, intention, time management, and learning experiences, to name a few [3, 4]. Learning tactics and strategies adopted by MOOC participants have been identified as key factors of success prediction [5–7]. Much research in traditional learning environments explored students' learning strategies [6, 8].

However, students' learning strategies in MOOCs are much less understood. MOOC platforms allow for recording trace data of the actual learners' behavior. However, such data are large, diverse, and complex to analyze. As a consequence, researchers have proposed a variety of methods that go beyond traditional statistics methods to unveil students' learning strategies [9, 10]. While the applied data analytic methods led to useful findings, the diversity of the adopted methods hindered the replication and generalization of the results. Little work has been done to compare how the applied approaches differ in terms of the tactics and strategies that they identify. This study explored how three analytic approaches – drawing from sequence, process, and network analytic techniques – could influence the detection of learning tactics and strategies.

# 2 Background

Research has emphasised the importance of using effective learning strategies as one of the key factors of successful learning. Learning strategy can be defined as "any thoughts, behaviors, beliefs or emotions that facilitate the acquisition, understanding, or later transfer of new knowledge and skills" [12, p. 727]. A closely related construct is the one of learning tactic, which can be defined as a sequence of actions that a student performs in relation to a given task within a learning session [12]. Defined in terms of tactics, learning strategies can be considered the regularity in the application of learning tactics or a pattern of how each student uses certain tactics [13]. Such patterns of tactic application evolve and become the characteristics of one's learning, which may be considered as aptitudes that could further predict the future behaviors [14].

Thanks to the large dataset of trace data on students' behavior, contemporary research aims to leverage these datasets to explore learning tactics and strategies by considering how these dynamic constructs unfold. In *network analytic approaches*, learning tactics and strategies are identified from networks built based on the co-occurrence of learning states or actions. These approaches were originally proposed for studying learning strategies as learning sequences [15]. The application of graph multiplicity measures, as commonly used in network science, has been then suggested to analyze the importance of individual events that contribute to student learning. For example, Siadaty et al. [16] applied this methodology to identify how technological interventions activated different processes of self-regulated learning. More recently, approaches suggest the use of *sequence analysis* techniques combined with unsupervised learning to detect learning tactics and strategies from trace data [9]. Similarly, learning tactics and strategies can be identified by analyzing the distribution of learning sequences [17].

Process-oriented data analysis approach emphasise the timing of the events. Malmberg et al. explored self-regulated learning strategies in a collaborative learning context by using a process mining technique [18]. Similarly, Matcha et al. [10] detected learning tactics and strategies from trace data by combining temporal analysis of the trace data (first-order Markov models) and clustering (Expectation-Maximization) [10]. Maldonado-Mahauad et al. [19] used a combination of process mining and clustering techniques to identify self-regulated learning strategies that different group of learners employed when interacting with the course contents (video-lectures and assessments).

Despite the interesting insights produced by these individual approaches, there has been limited research that explored how these three analytic approaches might have influenced the results. Hence, this paper aims to answer the following research question: How do different data analytics techniques proposed in the literature for the detection of learning tactics and strategies apply to the same dataset? That is, the paper compares approaches that emphasize sequence, network and process dimensions.

#### 3 Methods

#### 3.1 Data

The data used in this study was collected from the Introduction to Python course offered by the Pontificia Universidad católica de Chile on the Coursera MOOC platform in its two different editions. A total of 4,217 students registered their interest in the course. The course was in Spanish and was offered on demand (i.e. self-pace). In 8 weeks, the course covered six programming topics with 2-3 subtopics each. For each topic, the course offered a set of short video lectures with embedded questions (to provoke a simple recall of the concepts) and a set of reading materials. The students also had several theoretical exercises (11 quizzes) and practical exercises (13 exams). Among the quizzes and exams, 22items were graded and accumulated to calculate students final mark. At least 80 percent of these items had to be answered correctly to pass the course. The students were also offered the discussion board to discuss course topics. In this study, we considered only the trace data of those students who completed at least one assignment during the official course schedule between September 17<sup>th</sup> and November 4<sup>th</sup> 2018. As a result, 368 students were considered for the study. We coded the different learning actions captured in the trace data as described in Table 1.

**Table 1.** Coding of learning actions from the data trace

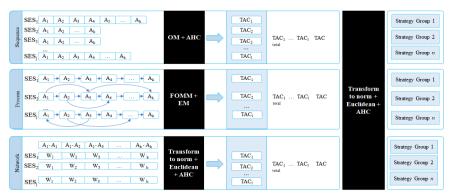
Events	Coded events	Description		
Video Lec-	lecture_start	Start the video lecture		
ture	lecture_complete	Complete the video lecture		
	in_video_quiz	Answer a quiz embedded in the video		
	In_video_quiz_correct	Correctly answer a quiz embedded in the video lecture		
	In_video_quiz_incorrect	Incorrectly answer a quiz embedded in the video lecture		
Reading	Supplement_complete	View the supplementary documents		
Theoretical	Quiz_start	Start a theoretical exercise		
exercises	Quiz_complete	Complete a theoretical exercise		
	Quiz	Theoretical exercise progress		
Practical	Exam_start	Start a practical exercise		
exercise	Exam_complete	Complete a practical exercise		
	Exam	Practical exercise progress		
	Exam_correct	Correctly solved a practical exercise		
	Exam_inccorect	Incorrectly solved a practical exercise		
	Code_execute	Command to execute the code		
Discussion	Discussion_question	Post a question to the discussion board		
	Discussion_answer	Post an answer to a question in a discussion board		
	Discussion_question_vote	Vote for a question		
	Discussion_answer_vote	Vote for an answer to a question		
	Discussion_answer_del_vote	Deleted a vote for an answer		
	Discussion_follow	Flag to follow a discussion		
	Discussion_unfollow	Flag to unfollow a discussion		

The resulting dataset for the analysis study contained the following items for each learning actions: the anonymous user ID, timestamp, type of learning action, and reference to course items. Each two consecutive learning sessions were separated

by at least 30 minutes of inactive time [20]. Due to the requirements of analytic methods to be applied, the outliers were excluded: extremely short sessions (one action in a session) and extremely long sessions (>95<sup>th</sup> percentile of actions per session).

#### 3.2 Methods

Fig. 1 illustrates the pipeline of the analytic methods used to extract learning tactics and strategies from the trace data following the three analytic approaches discussed in Section 2. The data were pre-processed based on the requirement of each analytic approach.



(\*SES: Learning Session; A: Learning Action; W: Weight of co-occurrence between two actions; FOMM: First Order Markov Model; EM: Expectation-Maximization; OM: Optimal Matching Score; AHC: Agglomerative Hierarchical Clustering; TAC: Learning Tactic)

Fig. 1. The pipeline of the analytic methods used in the study

**Sequential Dimension.** Following the work in [9], the TraMineR R package [21] was used to explore the sequential data. Learning actions were arranged chronologically and split into learning sessions. Sessions were encoded as learning sequences using a TraMineR's sequence representation format [21]. The optimal matching technique, with substitution costs based on transition rates, was used to compute the (dis)similarity of the sequences. Agglomerative hierarchical clustering based on Ward's algorithm was used to group learning sequences based on shared patterns of learning actions.

**Process Dimension.** The process dimension was explored by replicating the steps proposed in [10]. The pMineR R package was used to generate a process model of learning and compute the probability of state transitions by using the first-order Markov model (FOMM) [22]. The process model was formulated using timestamped learning events in each learning session. The Expectation-Maximization (EM) algorithm was used for clustering of learning sequences as it works well with the FOMM.

**Network Dimension.** The rENA R package for Epistemic Network Analysis (ENA) was used to compute the co-occurrence of learning actions in each learning session [23]. By generating a network using ENA, a matrix of co-occurrences of learning actions was created. The co-occurrence values in the matrix were normalized and subsequently used as an input to the agglomerative hierarchical clustering, based on Ward's algorithm. The Euclidean method was used to calculate the (dis)similarity.

The clusters of sequences (i.e., tactics) detected by each of the three data analytic approaches were then explored in terms of sequence length and event distributions. The

similarities between the three approaches were also calculated as proportions of learning sessions shared across the tactics detected by the three approaches.

To compute learning strategies, we used the results of cluster assignments of each of the three above approaches. Specifically, for each student, we computed the counts of each of the detected tactics and the total count of tactics. These counts were then normalized (i.e., reduced to the range of 0 to 1) and used as input to the agglomerative hierarchical clustering method. The computation of the (dis)similarity of students' tactic use was based on the Euclidean metric. The identified clusters were considered manifestations of the students' learning strategies (i.e., patterns of learning tactics). This was done for each of the three examined approaches. The identified learning strategies were explored based on how students applied the tactics according to the course topics. Furthermore, the association of the identified strategies and the final course marks was examined using Kruskal Wallis tests followed by pairwise Mann Whitney U tests.

# 4 Results

#### 4.1 Learning Tactics

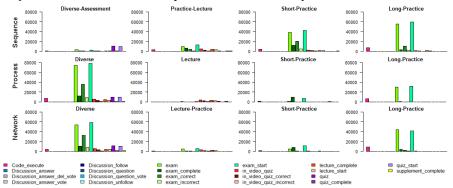
The results revealed that the three detection approaches identified four similar learning tactics. Fig. 2 presents the counts of learning actions in each tactic as identified with the three analytics approaches. Further details of the tactic characteristics are provided in the supplementary document (Tables 1-3)<sup>1</sup>.

**Sequence Approach.** The dendrogram suggested four clusters as the best result. The *Practice and Lecture-oriented* cluster (N=3134 sessions, 59.34%) was the largest and contained the shortest sequences (Mdn = 10 actions). The most dominant actions included those related to the exam activities, interaction with the video lecture, and quizzes embedded in the video. The *Diverse Assessment-oriented* (N=208 sessions, 3.94%) cluster was very small and contained long sessions ranging from 54 to 355 actions. This tactic often began by interacting with the video lectures, followed by doing the exam and ended by interacting with the quiz items. The *Short Practice-oriented* (N=1292 session, 24.47%) cluster included practical exams and code as the most dominant actions. Access to the video lectures was also prominent. The length of the sequences was moderate as compared to the other three tactics (Mdn = 93 actions). The *Long Practice-oriented* cluster (N=647 sessions, 12.25%) was relatively small exhibiting a pattern similar to the previous one (Short Practice-oriented). However, learning sequences were longer, ranging from 103 to 359 events (Mdn = 214).

**Process Approach.** Four tactics were identified with the process analytic approach as optimal. The *Diverse* cluster (N = 2000 sessions, 37.87%) varied in the number of actions in each session in the [3-359] range (Mdn=105). The main learning actions were related to exam activities, followed by quizzes, code execution, and interaction with lecture videos. The *Lecture-oriented* cluster (N = 1391 sessions, 26.34%) contained short sessions (Mdn = 7 actions). The most dominant actions included interaction with the video lectures and the quizzes embedded in the videos, followed by interaction with the quizzes that were part of the theoretical questionnaires. The *Short Practice-oriented* cluster (N = 772 sessions, 14.62%) consisted mostly of short sessions (Mdn = 8 actions)

<sup>&</sup>lt;sup>1</sup> Supplementary Document can be found at: https://bit.ly/2E4pFCu

that were generally of two types: i) short sessions of code executions and ii) longer sessions of completing an exam. The *Long Practice-oriented* cluster (N= 1118 sessions, 21.17%) mostly included actions related to the exam or code execution. Unlike the *Short-practice-oriented* tactic, completed exams were rarely observed in this tactic.



**Fig. 2.** The distribution of learning action counts across the tactics detected by the three analytic approaches

**Network Approach.** The dendrogram inspection suggested four clusters as optimal. Diverse-oriented (N = 1892 sessions, 35.83%) was similar to the Diverse tactic detected by the other two approaches; this tactic included a variety of actions, dominated by those related to exam and quiz related activities. However, the number of actions within a session was much higher compared to the Diverse tactic detected by the other two methods (Mdn = 93 actions). Lecture and Practice-oriented (N = 929 sessions, 17.59%) was the most dominant with exam-related actions and a small proportion of actions related to the lecture videos. However, when inspecting all the sequences, this cluster contained multiple short sessions of video lecture related actions often followed by long sequences of exam related actions. Unlike the Lecture and Exam-oriented tactic detected by the process analytic approach, the frequency of interactions with exam items outnumbered lecture-related actions, while quizzes-related actions were almost invisible. Short Practice-oriented (N=1776 sessions, 33.63%) was similar to the Short Practice-oriented tactic detected with the process approach. This tactic consisted of short learning sessions (Mdn = 7 actions). It was dominated by two types of sequences: i) short session of code executes, and ii) longer sessions of initiating and completing an exam. Long Practice-oriented (N= 684 sessions, 12.95%) contained longer sequences of action (Mdn = 126 actions). The most dominant learning actions were related to the exam or code execution. The proportion of initiated but not necessarily completed exams and continuing doing the exam was relatively high.

#### 4.2 Comparison of Detected Tactics

The *Diverse* tactic detected by the process and network approaches showed similar patterns; that is, it was composed of several different learning actions and diverse length of sequences. The most frequent action was interaction with the exam, followed by the interaction with quizzes. *Diverse-assessment-oriented*, as detected by the sequence approach, showed that the interactions with the quizzes were more frequent than the ex-

ams. Lecture and practice-oriented included events about actions related to video lectures and exams as the most dominant. Opposite to the other two approaches, the lecture related events outnumbered the exam focused events in the case of the process approach. Short Practice-oriented was defined by intense interaction with the exam items and code implementation. The median length of sequences of this tactic was smaller than of that of the Long Practice-oriented tactic. This tactic, as identified by the sequence approach, had the highest mean length of sequences and higher frequency of video lecture interactions than the same tactic detected by the other two approaches.

The sequence approach proved to be the best in distinguishing *Long Practice-oriented* as the one characterized by long sessions of exam interaction and code execution. The process and network approaches showed inconsistency in categorising based on the length of the sequences.

Similarity: 1861 Sessions (35.24%)		Process Analytic Approach (100%)					
		Diverse -Prac- tice	Lecture	Long-Practice	Short-Practice		
Sequence Analytic Approach	Diverse-Assessment	9.25	1.65	0	0		
	Lecture and Exam	22.4	98.35	58.94	85.36		
	Long-Practice	21.6	0	18.34	1.3		
Approach	Short-Practice	46.75	0	22.72	13.34		
Similarity: 1500 Sessions (28.40 %)		Network Analytic Approach (100%)					
		Diverse – Practice	Lecture and Practice	Long-Practice	Short-Practice		
Sequence Analytic Approach	Diverse-Assessment	10.84	0.32	0	0		
	Lecture and Exam	29.49	89.56	14.62	92.57		
	Long-Practice	14.64	2.26	49.71	0.51		
Approach	Short-Practice	45.03	7.86	35.67	6.93		
Similarity: 3526 Sessions (66.77%)		Network Analytic Approach (100%)					
		Diverse – Practice	Lecture and Practice	Long-Practice	Short-Practice		
Process Analytic Approach	Diverse	85.68	17.33	25.44	2.48		
	Lecture	10.94	78.26	0	25.73		
	Long-Practice	2.11	3.34	69.44	32.21		
Approach	C1 . D .:		1 00				

**Table 2.** The similarity of tactics detection based on three analytic approaches

Table 2 compares the results of the three analytic approaches based on cluster assignments of study sessions. The similarity was computed by calculating the proportion of learning sequences that were categorized as the same tactic. The sequence approach had 35% of overlap in session assignment with that of the process analytic approach, and 28% with that of the network approach. Almost 67% of sessions were categorized as representing the same tactics by the process and network analytic approaches. The *Lecture-oriented* tactic showed a high consistency among the three methods. About 98% of sessions labelled as the lecture-oriented tactic detected with the process analytic approach were also categorised as the same tactic in the sequence analytic approach. This high consistency might be a result of the high number of short learning sessions that coincide with interaction with lecture videos. The highest inconsistency among the approaches was for the *Short Practice-oriented* tactic.

The process and network analytic approaches categorised 3,526 (out of 5,281) sessions as the same tactics. We further explored the sequences that were grouped differently to examine how the approaches differ in grouping the sequences. One of the examples is SequenceID13745 that consisted of actions shifting between practical

exam\_start and exam\_progress. Execution of code was also observed during the exam progress, as shown in Fig. 3. This session consisted of 29 actions, which were inferred as representative of the *Long Practice-oriented* tactic by the process analytic approach. However, in case of the network analytic approach, the *Long-practice oriented* tactic had a higher median session length (Mdn=126), so that the considered sequence (SequenceID13745) was not qualified as an instance of the *Long Practice-oriented* tactic, but rather fitted in the *Short-Practice-oriented* tactic.

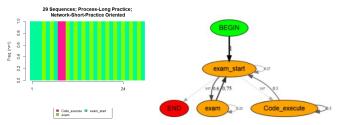


Fig. 3. The visualisation of sequenceID13745 and its first order Markov model

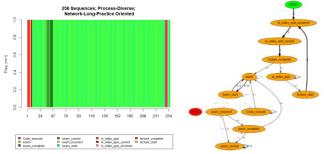


Fig. 4. The visualisation of sequenceID21601 and its process model

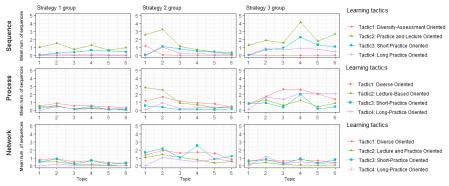
Another example of differences in the tactic detection is SequenceID21601 which contained 256 events. The sequence began by interacting with a quiz in a lecture video, followed by transitions between exam\_start, exam progress, a correct/incorrect exam answer, and exam complete; the command to execute the code was observed towards the end of the session, as presented in Fig. 4. The sequence and network analytic approaches associated this session with the *Long Practice-oriented* tactic. This is reasonable, since this sequence was relatively long, and the events showed dynamic transitions between the exam related actions. Meanwhile, the process analytic considered this sequence as an instance of the *Diverse* tactic. This is presumably because the sequence began by interacting with the video lecture. The Diverse tactic exhibited events about a variety of learning activities in a session.

### 4.3 Learning Strategy Groups

Learning strategies were identified as patterns of how students regulated the tactics according to the study topic. Detail characteristics of each strategy group are provided in the supplementary document (see footnote 1).

**Sequence Approach.** Three strategy groups were extracted based on how the students employed the tactics identified with the Sequence approach. Fig. 5 presents the

mean number of tactics employed according to the studied topic. Strategy Group 1 (N = 151 students, 41.03%) exhibited a low level of engagement. The dominant tactic was Lecture-oriented with short sessions. The mean number of sessions declined as the course topic progressed for all tactics except for the Short Practice-oriented tactic. The students who employed this strategy pattern had a high rate of failing the course (77.48%); their median course grade was 36.14 over 100, and the median number of passed graded items was 9 (out of 22). *Strategy Group 2* (N = 151 students, 41.03%) exhibited a high level of engagement when interacting with the first two topics by utilising the Lecture-oriented tactic. The Short and Long practice-oriented tactics increased when the course reached the second topic. However, the level of engagement dropped remarkably after completing the third topic. This strategy group had the highest failure rate (88.74%). The median of the completed graded items was four, and the median course grade was 18.04. Strategy Group 3 (N = 66 students, 17.94%) had the highest course grade (Mdn = 82.86/100), highest number of passed graded items (Mdn = 20 items), and the smallest failure rate (54.55%). Similar to the other strategy groups, the students frequently used the Lecture-oriented and Short practice-oriented tactics. Unlike the first two strategy groups, the mean number of sessions increased as the students moved to more difficult topics.



**Fig. 5.** Frequency of tactics used for each topic and for each strategy group as detected by the three analytic approaches

**Process Approach.** The mean number of employed tactics detected based on the process analytics approach according to the studied topic is presented in Fig. 5. *Strategy Group 1* (N = 215 students, 58.42%) exhibited a low engagement level. The mean number of sessions was consistently below one per study topic. The students who adopted this strategy had a high failing rate (82.79%); their median course grade was 29.33 over 100, and the number of passed graded item was 7 out of 22 items. *Strategy Group 2* (N = 89 students, 24.18%) included the students who were quite selective. The *Lecture-oriented* and *Diverse* tactics were dominant at the beginning of the course. The level of engagement dropped constantly from Topic 3 onwards. Despite putting a higher level of effort than *strategy group 1*, the students in this group passed less graded items (Mdn = 5), and had lower course grade (Mdn = 20.41). *Strategy Group 3* (N = 64 students; 17.39%) showed the highest passing rate (43.75%) and grades (Mdn = 82.71/100). Unlike the other strategy groups, the students in this group were consistently increasing

their engagement with the course topics. As the MOOC progressed and the topics became more challenging, this group put more effort and used diverse learning tactics, as shown by the high use of the *Practice-oriented* and *Diverse* tactics.

**Network Approach.** Fig. 5 shows three strategy groups with similar tactic enactment patterns. *Strategy Group 1* (N = 188 students, 51.09%) used multiple tactics but with a low frequency, and the frequency decreased as the course progressed. The group had a high failure rate (83.51%) with the median score of 27.29 (over 100), and passed, on average, 7 (out of 22) graded items. *Strategy Group 2* (N = 94 students, 25.54%) included the students who were the most active. They employed a variety of tactics to study each topic. The use of the *Diverse* and *Lecture-oriented* tactics slightly declined as the course progressed. There was some fluctuation in the use of the *Short Practice-oriented* tactic, especially during the fourth topic. The students with this strategy had the highest course score (Mdn = 56.95), and passed more graded items (Mdn = 15 items) than those following the other two strategies. *Strategy Group 3* (N = 86 students, 23.37%) had a similar pattern as the first one. Yet, the rate of students who failed was lower (74.42%), and the median grade was higher (Mdn = 37.04) than for strategy 1.

Association with performance. The strategy groups detected by using the sequence approach showed no significant association with the course grade, nor with the number of item passed (Table 3). However, we detected a significant association of the strategy and the potential of failing/passing the course. The pairwise comparison (Table 3) showed statistically significant associations among all the strategy groups and the potential of failing/passing the course. The effect sizes ranged from small to medium.

**Table 3.** Kruskal-Wallis (above) and pairwise comparison (below) of strategy groups with respect to performance

	Sequence	Process	Network
Course Grade	p = 0.125	p = 0.067	p = 0.14
Passed Graded Items	p = 0.082	p = 0.0004*	p = 0.01*
Passed Course	p = 0.046*	p = 0.0004*	p = 0.025*

Approach	Item	Strategy	Strategy	Z	р	r
Sequence	Passed Course	S1	S2	2.607	0.014*	0.150
		S1	S3	-3.401	<0.001*	0.231
		S2	S3	-5.613	<0.001*	0.381
Process	Passed Course	S1	S2	-0.160	0.88	0.009
		S1	S3	-4.401	<0.001*	0.263
		S2	<b>S</b> 3	-3.463	<0.001*	0.28
	Passed Graded Items	S1	S2	0.102	0.87	0.006
		S1	<b>S</b> 3	-7.206	<0.001*	0.431
		S2	<b>S</b> 3	-6.359	<0.001*	0.514
Network	Passed Course	S1	S2	-2.440	0.05	0.146
		S1	S3	-1.765	0.18	0.107
		S2	S3	0.516	0.57	0.039
	Passed	S1	S2	-4.323	<0.001*	0.258
	Graded	S1	S3	-2.762	0.05	0.167
	Items	S2	S3	1.613	0.059	0.121

Note: \* marks statistically significant differences

The strategy groups detected with the process analytic approach had no significant differences in course grades. A significant association was present between the strategy

groups and the number of passed graded items and the potential of failing/passing the course. Pairwise comparisons of strategy groups with respect to the completed performance items showed significant differences between strategy group 1 and 3 and groups 2 and 3. The effect sizes were medium except for the passed graded items between strategy groups 2 and 3 where the effect size was large (r=0.514).

The strategy groups identified with the network analytic approach had no significant difference on course grades. The strategy groups proved to differ significantly with respect to the number of items passed and the potential of failing/passing the course. Pairwise comparisons showed significant differences between strategy groups 1 and 2 on the number of passed graded items with the small effect sizes.

#### 4.4 Comparison of Detected Strategy Groups

Table 4 summarises the detected strategy groups along several dimensions related to the students' pattern of course engagement and academic achievement.

**Table 4.** Comparison of the strategy groups as detected by the three analytic approaches

	Sequence	Process	Network
Highly active and multiple tactics used	Strategy3	Strategy3	Strategy 2
Highly active at the beginning	Strategy2	Strategy2	-
Surface engagement	Strategy1	Strategy1	Strategy1,Strategy3

Highly active and multiple tactics used. These strategy groups reflect the deep learning approach as defined by Biggs (1987). The deep approach is characterised by high efforts, a variety of learning tactics used [7, 10], and associated with the high academic performance [4]. The students employed a variety of tactics when interacting with each topic. Based on the sequence approach, the most dominant tactic used was Lecture-oriented. Based on the process and network approaches, the dominant tactics were Diverse and Practice-oriented. Regardless of the tactic detection method, a similar pattern of interaction with the fourth course topic was observed – high enactment of the Short Practice-oriented tactic. This suggested that students might have been facing some challenges with the fourth topic that the instructor should consider when designing the next course iteration.

Highly active at the beginning. The sequence and process analytic approaches detected this similar pattern of tactic use, but not the network approach. The students were actively engaged during the first two topics, and then the effort significantly declined. The tactics employed during the first three topics showed that students were strategic in choosing tactics. The dominant tactics were Lecture-oriented and Diverse. This reflects the Strategic approach to learning [24], characterized by the aim of achieving high performance with the strategic choice of tactics [8, 24]. As the students faced more difficulty, their learning strategy shifted from strategic to the surface approach to learning. This suggested that some interventions are needed to maintain the level of students' engagement with the third topic. This group showed high engagement as compared to the Surface group, but the group missed to complete a few graded items.

Surface engagement. This group represented the surface approach to learning. As defined by Biggs (1987), students who follow this approach to learning employ surface

effort and have low academic performance [8, 24]. In our study, the students who followed this strategy group exhibited a low level of engagement and high failure rate.

None of the analytic approaches identified strategy groups that were predictive of performance. A significant association was found between the strategy group and the passed graded items for all cases. The process analytic approach proved the best in detecting strategy groups predictive of the passed graded items.

#### 5 Conclusions

**Summary.** The findings in this study showed that sequence, process, and network analytic approaches can be used to detect meaningful learning tactics from MOOC trace data. The three approaches resulted in tactics that were similar to some extent (Table 2). The highest similarity (67% of detected tactics) was found between the process and network approaches. As for strategy detection, the results of the network analytic approach differed from the other two approaches. The sequence and process analytic approaches resulted in similar strategy groups.

In general, we observed that sequences with similar learning actions were grouped in the same cluster. The length of the sequences affected the clustering in the sequence analytic approach. For example, short learning sessions were grouped into a single cluster (i.e. short diverse oriented) and this was the key distinguishing characteristic of this tactic group. In contrast, the process and network analytics were less based on the length of the sequences. Therefore, in the tactics detected using these two approaches the number of actions per learning sessions varied, ranging from two to hundred or more.

The proportion of learning sessions that belonged to each of the detected tactics impacted the learning strategy detection. The sequence approach detected one large tactic, i.e. *Short Practice-Lecture oriented*, showed that all strategy groups were dominated by this tactic. Furthermore, we found that all of the strategy groups exhibited a high frequency of using the *Short Practice and Lecture-oriented* tactics. This is unsurprising considering the course design that emphasized the use of video lectures and practice exercises.

**Implications.** The key finding of the study is that the choice of the data analytic approach for detection of learning tactics and strategies affects the results. Specifically, the three approaches emphasize different dimensions of learning tactics – sequential, process, and network. The differences in the underlying modelling of the three analytic approaches produced different data representations that are then fed to an unsupervised (i.e., clustering) machine learning algorithm. The properties of these underlying representations – sequence, process, and network – had direct implications on the computation of the similarities between individual sessions, and thus, the way how clusters were formed to detect learning tactics. Moreover, the choice of the underlying modelling approaches for tactics had a direct impact on the choice of clustering algorithm. For example, the process approach produced the data structure (i.e., adjacency matrix) that was not suitable for analysis with AHC; EM was used instead as also used in the literature [10]. AHC was more suited for the other two approaches, as commonly applied in the literature on similar tasks [9].

Based on the results of our findings, we cannot indicate which of the approaches is 'best'. Instead, the (dis)similarities in the results the three approaches produced and

interpretations of the (dis)similarities in this study can inform decisions of researchers and practitioners who work on the detection of learning tactics and strategies. Given that each of the three approaches used unsupervised machine learning at its core, it is also important that the interpretation of results should be done by considering a well-grounded educational learning theory and the learning context the data originate from. In our case, we offered examples that grounded in the theory of approaches to learning and the design of the MOOC used in the study. The use of these two sources demonstrated that all three approaches produced practically and theoretically meaningful learning tactics and strategies.

The differences in the learning strategies detected by the three approaches can directly be attributed to the differences in the modelling approaches used for the detection of learning tactics. This is due to the use of the identical methodology applied in the second step of the three detection approaches (see Fig. 1). Future research should investigate the extent to which changes in the modeling approaches in the second step will influence the results in the detection of learning strategies.

**Limitations.** Some limitations of this research must be highlighted. First, the detection of learning tactics and strategies relied primarily on trace data. Although limitations of self-reports are well document [12, 25], self-reports could add to the understanding of students' conditions, intention and motivation. Moreover, using multimodal techniques to capture the data could offer a fine-grained dataset. Second, some degree of subjectivity was evident in the selection of the number of clusters identified, even though the selection was informed by the information generated with the clustering technique (e.g., dendrogram in agglomerative hierarchical clustering) and further informed by the interpretability of the cluster solutions. Future research should explore approaches that can be used to produce a 'stable' number of clusters across different contexts.

**Acknowledgements.** This study has been partially funded with support from the European Commission through the LALA project (grant No. 586120-EPP-1-2017-1-ES-EPPKA2-CBHE-JP).

# References

- Zurita, G., Hasbun, B., Baloian, N., Jerez, O.: A Blended Learning Environment for Enhancing Meaningful Learning Using 21st Century Skills. (2015).
- Drachsler, H., Kalz, M.: The MOOC and learning analytics innovation cycle (MOLAC): A reflective summary of ongoing research and its challenges. J. Comput. Assist. Learn. 32, 281–290 (2016).
- Kizilcec, R.F., Pérez-Sanagustín, M., Maldonado, J.J.: Self-regulated learning strategies predict learner behavior and goal attainment in Massive Open Online Courses. Comput. Educ. 104, 18– 33 (2017).
- Broadbent, J., Poon, W.L.: Self-regulated learning strategies & academic achievement in online higher education learning environments: A systematic review. Internet High. Educ. 27, 1–13 (2015).
- 5. Winne, P.H.: How Software Technologies Can Improve Research on Learning and Bolster School Reform. Educ. Psychol. 41, 5–17 (2006). https://doi.org/10.1207/s15326985ep4101.
- Yip, M.C.W.: Differences in Learning and Study Strategies between High and Low Achieving University Students: A Hong Kong study. Educ. Psychol. 27, 597–606 (2007).

- Maldonado, J., Pérez-Sanagustín, M., Manuel Moreno-Marcos, P., Alario-Hoyos, C., Merino, P., Delgado-Kloos, C.: Predicting Learners' Success in a Self-paced MOOC Through Sequence Patterns of Self-regulated Learning: 13th European Conference on Technology Enhanced Learning, EC-TEL 2018, Leeds, UK, September 3-5, 2018, Proceedings. Presented at the (2018).
- 8. Chonkar, S.P., Ha, T.C., Chu, S.S.H., Ng, A.X., Lim, M.L.S., Ee, T.X., Ng, M.J., Tan, K.H.: The predominant learning approaches of medical students. BMC Med. Educ. 18, 1–8 (2018).
- 9. Jovanovic, J., Gasevic, D., Dawson, S., Pardo, A., Mirriahi, N.: Learning analytics to unveil learning strategies in a flipped classroom. Internet High. Educ. 33, 74–85 (2017).
- Matcha, W., Gašević, D., Uzir, N.A., Jovanović, J., Pardo, A.: Analytics of Learning Strategies: Associations with Academic Performance and Feedback. In: Proceedings of the 9th International Conference on Learning Analytics & Knowledge. pp. 461–470 (2019).
- 11. Weinstein, C.E., Husman, J., Dierking, D.R.: Self-regulation interventions with a focus on learning strategies. Handb. Self-Regulation. 22, 727–747 (2000).
- Hadwin, A.F., Nesbit, J.C., Jamieson-Noel, D., Code, J., Winne, P.H.: Examining trace data to explore self-regulated learning. Metacognition Learn. 2, 107–124 (2007).
- 13. Derry, S.J.: Putting learning strategies to work. Educ. Leadersh. 47, 4–10 (1989).
- Winne, P.H., Jamieson-Noel, D., Muis, K.: Methodological issues and advances in researching tactics, strategies, and self-regulated learning. (2002).
- Winne, P.H., Gupta, L., Nesbit, J.C.: Exploring Individual Differences in Studying Strategies Using Graph Theoretic Statistics. Alberta J. Educ. Res. 40, 177–93 (1994).
- Siadaty, M., Gašević, D., Hatala, M.: Associations between technological scaffolding and microlevel processes of self-regulated learning: A workplace study. Comput. Human Behav. 55, Part B, 1007–1019 (2016).
- Boroujeni, M.S., Dillenbourg, P.: Discovery and Temporal Analysis of Latent Study Patterns in MOOC Interaction Sequences. In: Proceedings of the 8th International Conference on Learning Analytics and Knowledge. pp. 206–215. ACM, New York, NY, USA (2018).
- Sobocinski, M., Malmberg, J., Järvelä, S.: Exploring temporal sequences of regulatory phases and associated interactions in low- and high-challenge collaborative learning sessions. Metacognition Learn. 12, 275–294 (2017).
- Maldonado-Mahauad, J., Pérez-Sanagustín, M., Kizilcec, R.F., Morales, N., Munoz-Gama, J.: Mining theory-based patterns from Big data: Identifying self-regulated learning strategies in Massive Open Online Courses. Comput. Human Behav. 80, 179–196 (2018).
- Kovanović, V., Gašević, D., Dawson, S., Joksimović, S., Baker, R.S., Hatala, M.: Penetrating the black box of time-on-task estimation. In: the Fifth International Conference on Learning Analytics And Knowledge. pp. 184–193 (2015).
- Gabadinho, A., Ritschard, G., Studer, M., Muller, N.S.: Mining sequence data in R with the TraMineR package: A user's guide. Dep. Econom. Lab. Demogr. Univ. Geneva, Switz. 1, 1–124 (2008).
- Gatta, R., Lenkowicz, J., Vallati, M., Stefanini, A.: pMineR: Processes Mining in Medicine, https://cran.r-project.org/package=pMineR, (2017).
- Shaffer, D.W., Collier, W., Ruis, A.R.: A tutorial on Epistemic Network Analysis: Analyzing the Structure of Connections in Cognitive, Social, and Interaction Data. J. Learn. Anal. 3, 9–45 (2016).
- 24. Biggs: Student Approaches to Learning and Studying. (1987).
- 25. Zhou, M., Winne, P.H.: Modeling academic achievement by self-reported versus traced goal orientation. Learn. Instr. 22, 413–419 (2012).