Understanding Learner Behavior Through Learning Design Informed Learning Analytics

Hao Shen
The University of Hong Kong
Hong Kong, China
shenhao@connect.hku.hk

Leming Liang
The University of Hong Kong
Hong Kong, China
lmliangc@hku.hk

Nancy Law*
The University of Hong Kong
Hong Kong, China
nlaw@hku.hk

Erik Hemberg
Massachusetts Institute of Technology
Boston, USA
hembergerik@csail.mit.edu

Una-May O'Reilly
Massachusetts Institute of Technology
Boston, USA
unamay@csail.mit.edu

ABSTRACT
A goal of learning analytics is to inform and improve learning design. Previous studies have attempted to interpret learners' clickstream data based on learning science theories. Many of these interpretations are made without reference to the specific learning designs of the courses being analyzed. Here, we report on a learning design informed analytics exploration of an introductory MOOC on Computer Science and Python programming. The learning resources (videos) and practice resources (short exercises and problem sets) are analyzed according to the knowledge types and cognitive process levels respectively, both based on a revised Bloom's Taxonomy. A heat map visualization of the access intensity on a learner resource access transition matrix and social network analysis are used to analyze learners' behavior with respect to the different resource categories. The results show distinctively different patterns of access between groups of students with different course performance and different academic backgrounds.

Author Keywords
Learning Design informed Learning Analytics; learning behavior; learner resource access transition matrix; learning trajectory; Social Network Analysis; MOOCs

CSS Concepts
•Applied computing~Education~E-learning•Human-centered computing~Visualization~Visualization application domains~Visual analytics

1 INTRODUCTION
A still open research question is how learning design and learning analytics can mutually inform each other to improve learning effectiveness through appropriate feedback to learners and learning designers. The ability to record learner behavior data at scale on MOOC platforms has provided opportunities for new findings through learning analytics. For example, studies have verified the doer effect in MOOC learning [13], i.e. learners' overall interactions with practice exercises are more strongly correlated to scores than watching videos or reading materials. Maiyuran et al. [21] further explored how the doer effect varies with respect to different topics and learners' academic backgrounds. Wang et al. [34] propose a novel approach to visualizing learner behavior through Detailed Access trajectories (DAT). The DAT offers both fine granularity and coarse-grained visualizations of learner behavior sequences based on their interactions with resources. However, these studies lack comprehensive learning design perspectives [35]. To bridge this research gap regarding learner behavior, we adopt the revised Bloom's taxonomy [14] as a theoretical framework to ground our analysis of MOOC learning design. We then investigate whether and how learner access behavior regarding different learning design elements relates to learner academic background and their course performance.

A learning trajectory is the path along which learning may proceed [33] and consists of a sequence of transitions. We define a transition as a sequence of **connected resource traversals** each starting **from a trigger** leading **to a target**. Transitions that deviate from a unidirectional, forward sequential access can be interpreted as *reviewing* events if the target is earlier than the trigger, or *resuming* events if it is later. For example, in the MOOC course MITx 6.00.1x, which is our research context, a student struggling with an exercise #49 "guess my number" may review video #48 "Floats and Fractions" to clarify a concept. If the learner returns to #49 to resolve the problem after reviewing #48, a resuming event occurs. We use the term learner behavior patterns to refer to **recurring learning resource access transitions generalized from the learner behavior data**. Learners' backgrounds may have an impact on their learning

L@S '20, August 12–14, 2020, Virtual Event, USA

ACM ISBN 978-1-4503-7951-9/20/08.
https://doi.org/10.1145/3386527.3405919

behavior, such as prior knowledge, prior educational experience and specialty [7; 32]. Finally, learning design refers to the process of constructing effective learning environments, resources and tasks to help learners achieve targeted learning outcomes [16; 23]. In this paper, we categorize the properties of the learning resources and focus upon the categories in our learning design analysis. We aim to address two research questions through learning design informed learning analytics:

1. Do learner behavior patterns reflect specific learning design features of the MOOC course?

2. Which properties of the learning resources are more likely to be triggers or targets of learners' review behavior? Are the associations different depending on learners' academic achievements and their background knowledge?

The contributions of this paper include:

- A representation of resource access transitions from trigger to target visualized as a 2-dimensional heat map.
- Analysis of learner resource access transitions and their deviations from the intended sequence in learning design.
- Learning resource categorization with the revised Bloom's taxonomy [14] framed with a learning design perspective.
- Investigation of how learners' review and resume patterns relate to the learning design and learners' background.

The paper is organized as follows. In section 2 we present the background. The course context and data are introduced in section 3. The method is described in section 4. Section 5 contains the experiments. Finally, conclusions and future work are in section 6.

2 BACKGROUND

The present study is based on research in the areas of learning design and learning analytics that tries to achieve the goal of learning design informed learning analytics.

2.1 Learning Design

Experienced teachers often have a good understanding of the background of the students and are thus able to offer differentiated instruction and support to different learners. Learning design is the top-down process through which the teachers (and other stakeholders) construct effective learning environments, resources and tasks to help learners achieve targeted learning outcomes [16; 23]. Learning design usually starts with the identification of learning outcomes. Its outcome is usually a specification of the knowledge and skills to be achieved. The learning design process itself is often guided by pedagogical considerations regarding the kinds and sequence of learning tasks and resources that would be appropriate. It also includes the design of and placement of assessment tasks as an integral part of the learning design.

For courses that are delivered completely online, such as MOOCs, the pedagogical repertoire that teachers can implement is limited compared to blended learning situations. MOOC designers often realize their learning design intentions and pedagogical activities in the form of videos, reading materials and exercises, due to the MOOC platform constraints on the type of content and interactions of learning resources [9]. Teachers make design decisions on (1) the intended sequence of learning resources that learners should go through by watching videos or reading textual materials, and (2) the sequence and placement of exercises and tasks to provide practice opportunities to learners, and to provide assessment feedback on the extent to which learners have achieved the intended learning outcomes at different points of the course.

2.2 Learning Analytics

Learning analytics attracts researchers and educational stakeholders who wish to understand and improve learning through bottom-up analysis of learner behavior and consequential feedback [28]. They collect and analyze mostly observational data collected during the learning processes, and seek insight on students' learning through data science and analysis [29]. Hence, learning analytics is an inter-disciplinary field where various analysis methods and techniques are implemented [1]. Centered on learning, the fields aim to understand learning behavior, monitor learning processes, predict learners' performance, suggest intervention [29] and revise learning design [25]. An important area of study in learning analytics is to understand and/or model learning behavior. For example, machine learning models can predict dropout to provide insights into the high dropout rate among MOOC participants [19]. These can also identify the time or specific points in the course where learners drop out [e.g. 12]. Visualization techniques provide an alternative approach to identify and interpret behavior patterns. For example, Shi et al. [26] developed VisMOOC, a tool to visualize the learning behavior and patterns extracted from clickstreams in MOOCs. Derived from historical data, these analyses often pinpoint problem areas but they lack or have only very weak connections to the original learning design.

There have also been efforts to transform MOOC learners' clickstream data into intentional cognitive actions. For example, Lei et al. [17 p.4] used an expert designed aggregation method to transform learners' clickstream data into action types such as "assessment, auxiliary, social interaction, navigation, and reading". Sinha et al. [31] similarly aggregated learners' online behavior and interpreted them into different cognitive categories using Lang's [15] limited capacity information processing cognitive framework. These categorizations do not refer to the learning design of the resources or tasks, and the validity of the categorization/interpretation has not been substantiated in either of these papers. Thus, while there are efforts to connect the interpretation of learning analytics results with literature in the learning sciences, gaps remain [24].

In MOOCs, learners do not necessarily follow the course resource sequence intended by the learning designer; they are free to pursue their own learning sequences [5] aka learning trajectories in the mathematics education literature [33]. A learning trajectory as originally proposed by Simon [30] is a "path by which learning might proceed", and it is also referred to as a "hypothetical learning trajectory" because the real path is unknown before learning actually takes place. With the emergence of more quantitative means of observing learning and learning analytics, an exciting intersection arises: learners' actual learning trajectories can be derived and compared to the hypothetical. For example, Davis et al. [6] identified video interaction patterns in learner behavior. Chen et al. [3] explored ways of visualizing learners' sequences in their access to different learning resources. Wang, et al. [34] introduced the *detailed access trajectory* and used them to identify several learning behavior patterns in a MOOC. However, these works stop short of addressing one major challenge in learning analytics, that of informing appropriate pedagogical action.

2.3 Learning Design informed Learning Analytics

There have been calls for combining learning analytics with learning sciences to overcome the learning challenges that arise [8]. Lockyer et al. [20] argue that learning design could serve to support this vision. By describing the pedagogical intents, and how these relate to the course components, learning designs can serve as the basis for the interpretation of learning analytics results. Bakharia et al. [2] further propose a conceptual framework which considers how learning design and learning analytics together empower teachers to make decisions on interventions. Guided by such a perspective, Shibani et al. [27] implemented a writing analytics tool that provides pedagogically meaningful analytics to teachers as well as learning feedback to learners, which significantly improved learners' performance. Furthermore, learning design can advance the application of learning analytics in authentic learning contexts, for example, there are some empirical studies in learning design informed learning analytics [4; 11; 22]. This paper extends these studies by presenting a study that explores and interprets learners' learning trajectories using learning design informed learning analytics.

3 CONTEXT OF RESEARCH AND DATA SOURCES

We use data from the MOOC "*Introduction to Computer Science and Programming Using Python*" (MITx 6.00.1x, 2017 spring) delivered on the EdX platform. In exercises and discussions with an instructor of the course we documented the design intent. The course comprised of 13 units (topics) taught over seven weeks, with two sessions per week. In each unit, learners need to watch lecture videos and complete short exercises (referred to as finger exercises in this course). The MOOC has no written content as an additional type of learning resource. There are 233 learning resources that can be accessed by learners (74 videos, 96 finger exercises, 40 problem sets and 23 examination questions). The videos introduce programming concepts and demonstrate coding examples of the application of these concepts. Finger exercises are made up of multiple- choice questions and/or short coding exercises that test the retention and application of the concepts delivered in the videos. Therefore, by design, the finger exercises are placed following relevant videos. At the end of each unit, a problem set with several problems to solve by submitting source code is used to test learners' abilities to make use of the major concepts and skills in the unit.

The data consist of learner generated data and course design data. We select a learner sample (n=1,561) out of 69,420 enrolled learners. This sample comprises all learners who completed the course with a passing score and paid 50 USD for a certificate, as well as responded to an optional pre-class survey. The learner survey findings show that those learners who have successfully completed the course came from different academic backgrounds (see Table 1).

By exploring the learning trajectories of these learners who have persisted through the course and achieved the targeted learning outcomes, we can investigate the nature of the resources and tasks that they find difficult, and whether there appears to be differences in learners' learning trajectories that are correlated with their academic backgrounds and course achievements. An added advantage of analyzing learners who were certificated is the availability of the learners' midterm and final exam scores.

Student profile	Categorization	Criteria	N
Education background	Higher education	Doctoral/Master's/Bachelor's/Associate degree	1057
	Non higher education	Junior secondary/junior high/middle school, elementary/primary school, no formal education, and unknown	354
Previous programming experience	Know programming	Know Python, know other languages, veteran	935
	No experience	No experience and unknown	625
Course grades	above 90	>=90	932
	between 80 and 90	>=80 and <90	264
	below 80	<80	365

Table 1. Learner subgroups by academic achievements and backgrounds. The student profile refers to learner academic backgrounds and achievements; the categorization column shows the specific categorizations of student profiles; the criteria column shows the standards used for making the classification decisions.

4 METHOD

To achieve the goal of conducting learning design informed learning analytics, a variety of methods are used to analyze course learning design and learner behavior. For analysis from a learning design perspective, we categorize the learning resources and tasks according to their knowledge types and cognitive demands respectively. We also use computational techniques to analyze and visualize the learners' access patterns through the resources and tasks. This culminates in a learner resource transition matrix that can be visualized with a heat map. Finally, using social network analysis, we examine similarities and differences in access patterns across learner groups with different academic backgrounds. The methods are next described in detail.

4.1 Content Analysis of Learning Resources

We hypothesize that learner review behavior is related to the nature of the content of learning resources, which is an important dimension in learning design. The review behavior reflects that the learner recognizes some learning difficulty in topics addressed in a previously visited learning resource that needs to be resolved in order to proceed to the next resource in the sequence. We hypothesize that the cognitive features of those learning resources may associate with learners' review behavior. We adopt Krathwohl's [14] revised Bloom's Taxonomy for learning, teaching and assessment in the analysis of the course content. The taxonomy comprises two dimensions: knowledge type and cognitive process. The knowledge type dimension categorizes the knowledge that is to be learnt from the resources into four types: factual, conceptual, procedural and metacognitive. Factual knowledge comprises the basic elements such as terminology and symbols in a discipline that learners must know to engage in activities in that discipline. Conceptual knowledge concerns relationships among basic elements in a discipline, such as classifications, principles, generalizations, theories, and models. Procedural knowledge is knowledge needed to conduct tasks. It includes methods of inquiry, discipline-specific skills and techniques, algorithms, and criteria for selection of procedures, etc. Metacognitive knowledge is knowledge about cognitive tasks, strategic knowledge and self-knowledge, which are particularly important when planning and solving problems. The cognitive dimension of Krathwohl's [14] taxonomy closely followed the original Bloom's taxonomy and comprises six verbs that describe different levels of cognitive demands in task performance: remember, understand, apply, analyze, evaluate and create.

To document the course design of MITx 6.00.1x, we conducted content analysis on the video lectures and finger exercises using the revised Bloom's taxonomy. The videos were designed to be instructional, so we reviewed the videos and categorized their content based on the knowledge type dimension. We applied the cognitive dimension to the analysis of the finger exercises and problem sets. Two researchers independently coded all of the resources. The intercoder reliability was found to be 0.857 ($P < .001$) using Cohen's Kappa measurement.

4.2 Learner Resource Access Transition Matrix

As mentioned earlier, we are interested in understanding learners' learning trajectory as represented by their resource access transitions in the learning process. The resource access sequence patterns of 1,561 learners through the 233 resources is difficult to manually analyze. We interpret the videos, exercises and problem sets presented to the MOOC learners as an explicit, linear, designed access sequence. In reality, learners may decide to deviate from the designed sequence if they find that sequence to be unsatisfactory for them. Thus, we are interested in identifying patterns of transition that deviate from the intended linear sequence, and to examine whether the high frequency "deviations" are connected to the type of learning resource and/or learner academic background.

Here we provide an explanation on how to interpret the learning trajectory from the transition matrix, using our research context, the MITx 6.00.1x MOOC course. A learner usually starts the course by viewing video #0 "Introduction", then moves to the next video #1 "Knowledge", followed by #2 "Exercise 1". Thus, if the learner stops learning at this point, his/her learning trajectory would be the sequence (#0, #1, #2), which includes two transitions (#0, #1) and (#1, #2). However, a learner's trajectory does not always follow a linear sequence of transitions in a natural order. It is known that MOOC learners can decide which resource and when to learn, i.e. they tend to design their own learning trajectories [10]. We observe that many learners went back to view #0 after watching video #17 "Bindings", following which transition they returned to #17 afterwards. In this case, the transition (#17, #0) is a review behavior and transition (#0, #17) is thus a resume behavior.

Using a matrix in which the columns (i.e. x-axis) represent destination resources and the rows (i.e. y-axis) represent departure resources, we could mark the frequency of all the "from-to" pairs onto the matrix, which we refer to as the "learner resource access transition matrix", or simply the *learner transition matrix*. We can also visualize the frequencies of the occurrence of each "from-to" pair on the matrix in the form of a heat map (see Figure 1).

We can formally define learning trajectories based on an access graph $G = (E, N)$. N is the vector of resource nodes, $N = \{n_0, \dots n_m\}$, $|N|$ is the total number of resources. E is the edges (traversals) between the resource nodes $(n_i \ , n_j \ )$. A is the adjacency matrix $A^{m \times m}$ where element $a_{ij}, 0 \leq a_{ij}$ indicates the number of traversals from node n_i to n_j. A learning trajectory is the ordered sequence (vector) of nodes $T = [n_i^0, \dots, n_j^k], 0 \leq i \leq m, 0 \leq j \leq m, 0 \leq k$. In the learning trajectory we classify each ($y_{\text{coordinate}}$, $x_{\text{coordinate}}$) pair $(n_i^t, n_j^{t+1}), 0 \leq t \leq |T|$ as a review if $j < i$ and resume if $j > i$ and $n_j^{t+1} \in [n_i^0, \dots, n_j^t]$. The case where the resume is back to the previous trigger is when $n_j^{t+1} = n_i^{t-1}$. This can describe both individual learning trajectories, as well as aggregate trajectories, where $A = \sum_{i=0}^{I} A_i$, $0 < I$ is the number of learners and A_i is the adjacency matrix for each learner. With this definition we can also define blocks of revisions and resumes as $a_{ij} > \alpha, a_{ij-k} > \beta, a_{ji-l,} 0 \leq j, 0 < l \leq i, 0 < \alpha, \beta$.

4.3 Social Network Analysis

We use SNA to shed light on the relationship between learners' access behavior and the nature of learning resources, because SNA enables us to quantify the importance of resources with particular types of learning resources. There are two key elements in SNA: nodes (aka vertex) and edges (aka ties) connecting the nodes. In this study, the nodes represent the learning resources in 6.00.1x, and the directional edges between two resources indicates learners' traversal from one resource to another. Then we use *degree* to measure the importance of a particular resource (node) in the access transition of learners. Degree refers to the number of edges connected to a particular node. SNA methods assume that a resource with a higher degree is more important in the network. Edges between resources are directional, so the overall degree can be divided into two types: *in-degree* and *out-degree*. In-degree represents the number of connections going *to* a resource, while out-degree shows the number of connections leaving *from* a resource. The number of different resources that are triggered for review from a particular resource is denoted as its out-degree, while the number of different resources that were triggered to review it is denoted as its in-degree. We assume that (1) a resource with high in-degree serves as a foundation (or obstacle) to the learning of other resources because many learners refer back to this resource, and vice versa; (2) a resource with a high out-degree is more likely to serve as a trigger for learners to review other resources, and vice versa.

We use Gephi 0.9.2 to generate the in-degree and out-degree for each learning resource, including videos, short exercises and problem sets. Then we use one-way ANOVA to analyze the correlation between the frequency of in/out-degree and learner background. The result can provide insight in two aspects: (1) whether the types of learning resource relate to learners' access transition, and (2) whether learners' access transition among different resources relate to learner background and academic achievements.

5 RESULTS & DISCUSSION

In this section, we first report on the knowledge types and cognitive process levels of the designed learning resources according to the revised Bloom's Taxonomy. Next, we use the learner transition matrix, social network analysis and one-way ANOVA methods to investigate the research questions: 1) Do the learner behavior patterns reflect specific features in the learning design of the MOOC course? And 2) Which types of learning resources are more likely to be a trigger/target of learners' review behavior? Are there differences dependent on learners' backgrounds?

5.1 Knowledge types and cognitive process levels of designed learning resources

Content analysis of the video lectures using the revised Bloom's taxonomy found three knowledge types: factual, conceptual and procedural knowledge, and these three types accounted for the majority (66/74) of the videos in the course. Metacognitive knowledge was not identified in any of the videos. The remaining 8 videos cannot be clearly categorized into any of the four knowledge types as these were either overviews of a topic (6) or a summary (2). We thus categorized these video into a separate category.

The finger exercises are tasks that required learners to practice applying what they have learnt from the videos. So we categorize them using the cognitive process categories. We found that only the three lowest level cognitive processes can be found in all 96 of the finger exercises: remember (14), understand (50), and apply (32). For the problem sets, we found it impossible to consistently categorize the problem sets into any one of the six cognitive processes as these are more complex problems that require multiple stages of problem solving, involving combinations of the six cognitive processes. Hence, we decided not to categorize the problem sets and simply retain the label "problem set".

5.2 Do the learner behavior patterns reflect specific features in the learning design of the MOOC course?

The learner transition matrix method (Section 4.2) generated a heat map to visualize the density distribution of resource transitions (see Figure 2). The color indicates the frequency for each transition pair. For instance, a dark color on the (y, x) coordinate (10,1) indicates a high frequency of learner access from resource n_{10} to n_1. We have super-imposed several horizontal and vertical reference lines to mark the boundaries between units and sessions in MITx 6.00.1x.

Resource type	N	Knowledge type/Cognitive process level	N
Video	74	Factual	12
		Conceptual	25
		Procedural	29
		Overview and Summary	8
Finger Exercises	96	Remember	14
		Understand	50
		Apply	32
Problem Set	40	Complex performance	40
Exam	23	-	-
Grand Total			233

Table 2. Knowledge types and cognitive processes of the resources.

Observations from the heat map visualizations of the learner transition matrix

In the heat map we can identify a diagonal line as the designed access sequence. A prominent line running from top-left to bottom-right of the figure is shown in Figure 2. The actual diagonal is a white line which represents a value of zero. This is expected as transitions are defined as traveling from one resource to another resource. The dark line immediately above the white diagonal shows a continuous sequence of accesses from resource n_i to resource n_{i+1}, which represents the learning trajectory as intended by the course designers. In addition, the coordinates on both sides of the diagonal appear to cluster into blocks. A block contains recurring traversal events happening within a fixed range. The sizes of the blocks differ. With the help of the superposed session/week reference line, we can see that many of the blocks align with the session and week boundaries in the course.

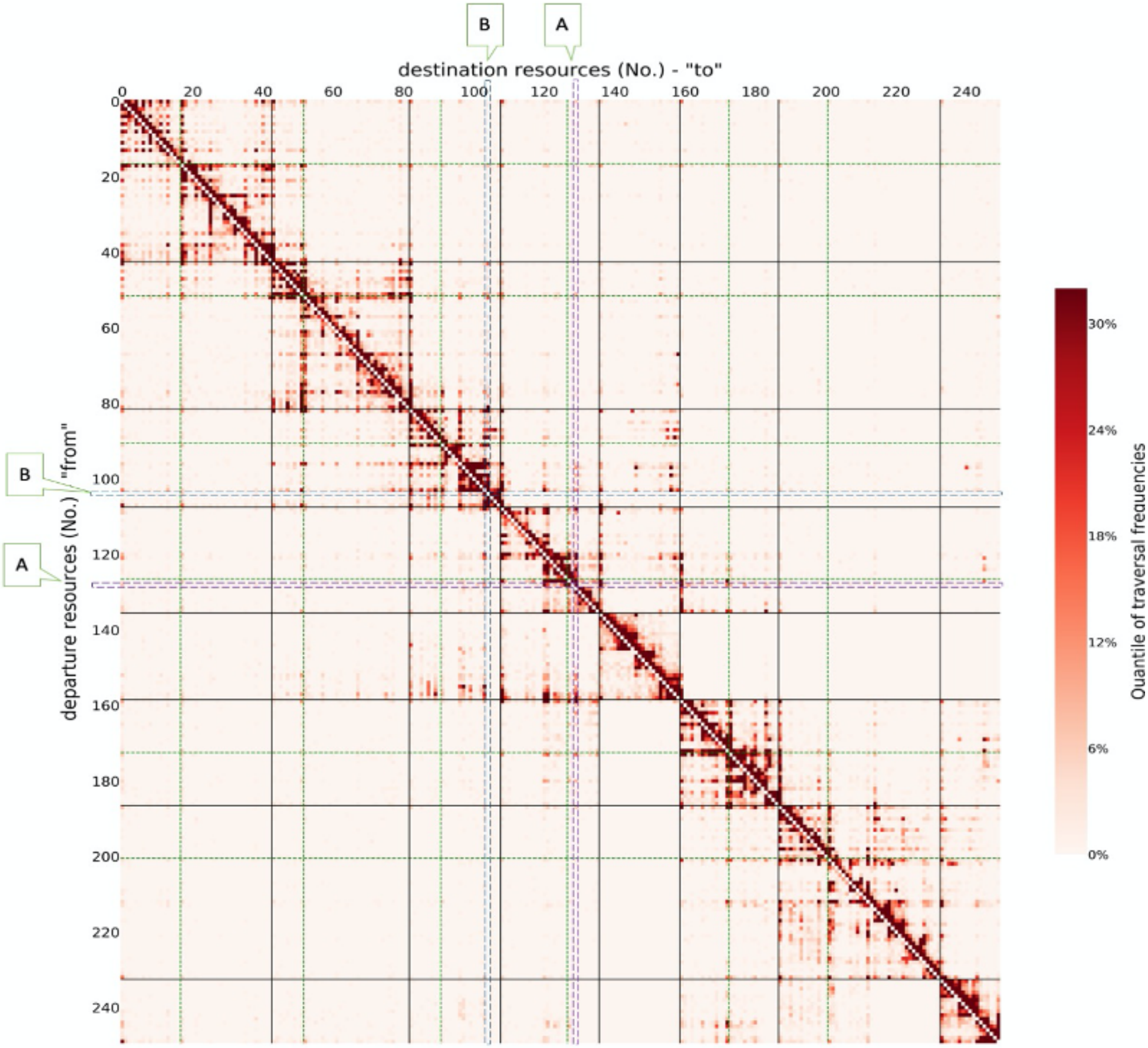

Figure 1. The heat map visualization of an aggregate learner resource access transition matrix for MITx 6.00.1x, where the y axis represents the trigger(from) resources, the x axis represents the target(to) resources and the color intensity of the elements corresponds to the frequencies of transitions between trigger resources and target resources. The solid lines indicate the dividers for each week unit of the course, while dotted lines indicate within-unit sessions. "A" is the label of resource #128 "Exercise 3" and "B" stands for resource #104 "Problem 1: Is the Word Guessed", both of which serve as examples in Section 5.3.2 of this paper.

Interpretations from the heat map visualizations of the learner transition matrix

Review-resume events: interpreting the "mirror image" around the diagonal. The diagonal separates the matrix into two parts. As explained in section 4.2, all traversal events below the diagonal represent reviewing events, because the "from" (y-axis) resource id numbers are larger than that of the "to" (x-axis) resources id numbers. Similarly, the marked traversals above the diagonal stand for resuming/catching-up events. Drawing on this interpretation, we see that at a global level, the reviewing and resuming events almost mirror each other, centering about the diagonal. It appears that reviewing and resuming events emerge in pairs, which is reasonable since learners are mostly likely to resume to the trigger resource after reviewing and resolving the problem.

The blocks indicating clustering of review-resume events within and across sessions. The blocks located along the diagonal indicate high interdependence in terms of the targeted knowledge and skills outcomes within the block. We believe this interpretation is credible as during documentation of the course, we observe the topics within

Resource type	Overall course score			Highest academic qualification			Knowledge of programming		
	Above 90	Below 80	Sig.	Higher Ed	Non- HE	Sig.	Know prog.	No prog.	Sig.
Video (N=74)	1.563	1.247	***	1.471	1.412	***	1.462	1.440	No sig.
Exercise (N=96)	1.292	1.005	***	1.203	1.164	***	1.213	1.162	***
Problem set (N=40)	1.024	0.731	***	0.932	0.923	No sig.	0.940	0.910	***
Midterm exam (N=23)	1.078	0.858	***	1.011	1.016	No sig.	1,017	1.007	**

Table 3. The per-learner frequencies of access by different learner groups to the different types of resources. For the learners grouped by the course score, academic qualification and knowledge of programming (i.e. the column titles except the first column), the number of their access frequencies of resources per learner in different resources types (as shown in the first column) are listed in the cells. The level of significance is reported comparing the difference of per-learner frequency between each pair in the subgroups.

the same session are carefully selected and likely to be strongly related to each other. Except for session/week blocks, some large but lighter colored blocks as well as some smaller dark blocks are also visible. The large lighter blocks reflect the review-resume events across sessions/weeks and may indicate interdependence in content across topics in different (designed) course sessions/weeks. For the smaller but dark blocks, they indicate similarities across the connections of sub-session topics and imply that there are more differentiated content in terms of subject-matter knowledge within a session.

5.3 Which types of learning resources are more likely to be a trigger/target of learners' review behavior? Are there differences dependent on learners' backgrounds?

In this section, we investigate whether there are interactions in the relationships between learners' academic backgrounds and their learning trajectory patterns. First, we examine the access frequency to the different types of learning resources based on learners' backgrounds. Then we focus on the in-degree and out-degree—i.e. the frequencies that a learning resource served as a target and a trigger respectively—for different types of learning resources.

5.3.1 Do the access frequencies of different resource types differ among learners with different backgrounds?

Table 3 summarizes the per-learner frequencies of access and the results of one-way ANOVA to compare the mean frequencies across each pair of subgroups. Among this sample of learners who have earned certificates, we can see significant differences among these groups. The learners who received a higher score, possessed a higher academic qualification or with more programming knowledge experience, all use course resources more frequently. Moreover, the difference is particularly prominent with regard to accessing videos.

5.3.2 Across three groupings of learners, do the in/out-degree frequency between different types of resources differ in the same grouping?

We investigate the in-degree and out-degree separately. We expect that for resources that are frequently accessed, the number of "visits from" (out-degree) and the number of "visits to" (in-degree) the same resource is similar. However, the number of different resources to review "triggered" by the departure resources is observed to be different from the number of resources that it served as the "target" resource for. For example, resources #128 "Exercise 3" (labeled as "A") of session 8 and #104 "Problem 1: Is the Word Guessed" (labeled as "B") of session 6 on the heat map are both frequently accessed. However, #128 triggered reviews of 90 resources as destination (out-degree=90), but only served as a review target from 39 resources (in-degree=39). The transitions of #104 are the opposite. It triggered reviews of only 44 resources as destination (out-degree=44), but served as a review target from 80 resources (in-degree=80).

We can describe the above differences in transition patterns of the resources using the Social Network Analysis (SNA). The number of different resources that are triggered for review from a particular resource is denoted as its out-degree, while the number of different resources that were triggered to review it is denoted as its in-degree. Further, we conducted a one-way ANOVA to compare the means of the in-degrees and out-degrees for different resource types for each category of learner academic background/achievement.

Observations from the comparisons of in/out-degree index between different types of resources in the same learner grouping background or academic achievement across three groupings of learners

The results in Table 4 reveal that there is a statistically significant difference across each pair of learner background. The most prominent difference is observed across learners with different course scores. For learners whose scores were below 80, there is no statistical difference in the in-degrees and out-degrees for the different types of resources (videos, finger exercises, problem sets). On the other hand, for learners whose scores were above 90, the in-degree of videos are higher than all other resources. The videos also had significantly higher out-degrees than all the other resource types, and the problem sets also had statistically higher out-degrees than the finger exercises for the higher performers.

The other groupings of learners were based on their backgrounds. When comparing the in-degree and out-degree differences of learners' access across the different learning resources, there is essentially no difference between learners who had prior programming knowledge compared to those who did not.

		Above 90	Below 80	Higher Edu	Non-Higher Edu	Know programming	No experience
Resource type	In-d	Video > Others	No significant differences	Video > Others Problem set > Finger exercise > Exam	Video > Others	Video > Others	Video > Others Problem set > Finger exercise
Resource type	Out-d	Video > Others Problem set > Finger exercise	No significant differences	Video > Others Problem set > Exam > Finger exercise	No significant differences	Video > Others Problem set > Finger exercise	Video > Others Problem set > Finger exercise

Table 4. Across three groupings of learners, comparing in/out-degree indices between different types of resources in the same learner grouping background or academic achievement. Prominent difference is observed across students with different course scores, while no significant difference is observed between students who had prior programming knowledge and those who did not. The greater-than sign indicates the mean in/out-degree for a certain resource group is greater than another, as the results of post-hoc tests if one-way ANOVA reports significance.

		Above 90	Below 80	Higher Edu.	Non-Higher Edu.	Know programming	No experience
Knowledge type	In-d	Overview and summary > Others	No significant differences	Overview and summary > Others	No significant differences	Overview and summary > Others	Overview and summary > Others
Knowledge type	Out-d	Overview and summary > Others	No significant differences	Overview and summary > Others	No significant differences	Overview and summary > Others	Overview and summary > Others
Cognitive process level	In-d	No significant differences	No significant differences	Analyze > understand	No significant differences	No significant differences	Analyze > understand
Cognitive process level	Out-d	Complex performance> remember	Remember < others	Complex performance > understand Complex performance > remember	No significant differences	Complex performance > understand Complex performance > remember	Complex performance > understand Complex performance > remember

Table 5. Comparing in/out-degree indices between different knowledge types of video and cognitive process levels of practice tasks for different learner backgrounds and academic achievements. The overview/summary videos had statistically higher values than all other types of videos. The learners with higher scores had the highest in- and out-degrees for the overview/summary videos, while their counterparts did not display any differentiation across the knowledge types. The greater-than sign indicates the mean in/out-degree for a certain resource group is greater than another, as the result of post-hoc test if one-way ANOVA reports significance.

On the other hand, there were some differences between learners depending on whether they already possessed any higher academic qualifications. Both groups of learners had significantly higher in-degrees for videos. However, there is a difference in the out-degree between the two groups. Those who had no higher degree qualifications had no out-degree differences across the different resource types, which is similar to learners scoring below 80.

Interpretations from the comparisons of in/out-degree indices between different types of resources in the same learner background grouping or academic achievement across three groupings of learners

Assuming that the high performers had better learning strategies, one interpretation of this result is that it indicates that the higher performers are more capable of using videos to seek clarifications and better understanding when they encounter difficulties. In terms of triggering learners' realization that they were encountering difficulties that required review to seek better understanding, the higher performers were more capable of identifying their own problems of understanding when receiving instruction from the videos and were also more able to identify complexities in understanding when working through the problem sets. While the learner background in programming knowledge did not affect their resource access transition behavior, general academic background did significantly influence the likelihood of them seeking better understanding through reviewing videos.

5.3.3 Do in/out-degree indices differ between the knowledge types of video and cognitive process levels of practice tasks based on learners' academic backgrounds and their course achievements?

Next, we proceeded to explore whether the average in-degree and out-degree frequencies of the videos and finger exercises show any statistical difference across knowledge types and the cognitive process levels of these two types of resources respectively for different learner groupings.

Observations from the comparisons of in/out-degree indices across different knowledge types of videos and cognitive process categories of practice tasks for different learner backgrounds and academic achievements

The results are shown in Table 5. For any knowledge type of the videos, there is no differentiation between learners whether they had prior programming experience or not. For both in- and out-degrees, the overview/summary videos had statistically higher values than all other types of videos. However, there is a statistical difference in both in- and out-degrees pertaining to different video knowledge types for both of the other two learner groupings. Both the higher performers and those with higher academic qualifications had the highest in- and out-degrees for the overview/ summary videos. On the other hand, for both the learners achieving a score below 80 and those without a higher academic qualification, their resource access behavior with regard to the videos did not display any differentiation across the knowledge types.

The lower part of Table 5 reports on whether there are different patterns of access by the different learner subgroups with regard to the cognitive process levels of the finger exercises and problem sets. In terms of the mean in-degree variations across different cognitive levels, there is no significant difference across any of the levels for both high performers and those with a score below 80. There is also no significant difference for learners without higher academic qualifications and those with programming experience. On the other hand, the complex performance tasks (i.e. problem sets) had significantly higher in-degrees for those with higher academic qualification and those without programming experience.

The one-way ANOVA results on the out-degree for the different cognitive process tasks reveal further differences. For learners scoring above 90, those with higher academic qualification, and for learners with or without programming experience, these groups of learners all have average out-degrees that are highest for complex performance tasks. On the other hand, there is no significant difference in the out-degree for learners with no higher academic qualification, and for those scoring below 80, there was only negative discrimination for the lowest cognitive level of finger exercise: remember.

Interpretations from the comparisons of in/out-degree indices between different knowledge types of video and cognitive process levels of practice tasks for different learner backgrounds and academic achievements

Maybe the learners achieving lower scores and those with lower academic qualification were not able to pick up the cues embedded in the overview/summary videos to guide their study strategy. The learners with higher academic qualification and those without programming experience could perceive a greater value to improving their understanding by reviewing the problem sets more than the finger exercises.

The problem sets are cognitively more demanding, requiring learners to analyze, evaluate and create programming solutions. For learners scoring above 90 and those with higher academic qualification, irrespective of whether they had any programming experience, they may be more sensitive to the different kinds of difficulties of encountered in the process of working on the problems, which triggered a more diverse set of target resources for review. The learners with lower scores or lower academic qualification maybe less aware of the full range of knowledge and skills that are needed for the complex performance required in solving the problem sets.

6 CONCLUSIONS & FUTURE WORK

In this study we follow a learning design informed learning analytics approach to understand learner behavior, more specifically the learner resource access trajectory. Previous studies investigating learner behavior paid little attention to the specific learning design features of a course in analyzing and understanding learner behavior. To bridge this gap, we categorize the learning resources in a MOOC course according to the revised Bloom’s taxonomy as the focal learning design features and try to find out the connection between learner behavior and learning design.

We constructed a learner resource access transition matrix to visualize how learners travel from one resource to another. The visualization shows that learners’ reviewing and resuming events take place more frequently within the same session, while some of them take place across different sessions. These findings indicate that learners’ access to the learning resources largely follow the intended learning trajectory as intended by the course designers. We further use the in-degree and out-degree indices of each resource computed using SNA to reflect the importance of a resource as a target or a trigger. Videos are more likely to serve as triggers or targets in off-sequence transitions for learners with higher scores or better academic backgrounds. These “better” performers are also more able to show discrimination in their access patterns for resources of different knowledge types and cognitive process levels. This is somewhat consistent with the previous study showing strong doer effects on problem set scores while video watching also matters on final scores [21]; this may also indicate “better” performers are able to utilize better learning strategies, in line with Littlejohn et al. [18]’s findings where self-regulated learners tend to adopt open and flexible task strategies. In this study, we extend the knowledge about how video watching matters based on access transition data and examined the transition behavior from the perspective of learning design based on cognitive categorization of the learning resources, and further explored the learning design connection in relation to learner background.

As with any research, there are a few limitations influencing the results and conclusions in this study. The research context is constrained within a single MOOC course, where we have only preliminarily adopted the Bloom’s taxonomy to document the learning design features of the course. More aspects and principles of learning design are need to be included in future learning design informed learning analytics studies.

In our future work, we plan to further investigate the intervention strategies for lower performers informed by better performers' learning trajectories. One possible implication of these findings is to provide different interventions for learners with different academic backgrounds and course achievements. We need to conduct a more fine-grained analysis on the learning design of the resources to gain a deeper understanding of learners' difficulties and provide customized pointers to resources to guide learners who may be less able to perceive their own learning difficulty.

ACKNOWLEDGMENTS

The research reported in this paper is fund by the Innovation and Technology Commission of the Hong Kong SAR Government, under the Innovative Technology Fund, grant number ITS/388/17FP.

REFERENCES

[1] Avella, J.T., Kebritchi, M., Nunn, S.G., and Kanai, T., 2016. Learning analytics methods, benefits, and challenges in higher education: A systematic literature review. Online Learning 20, 2, 13-29.

[2] Bakharia, A., Corrin, L., De Barba, P., Kennedy, G., Gašević, D., Mulder, R., Williams, D., Dawson, S., and Lockyer, L., 2016. A conceptual framework linking learning design with learning analytics. In Proceedings of the Proceedings of the Sixth International Conference on Learning Analytics & Knowledge - LAK '16 (2016), 329-338. DOI= http://dx.doi.org/10.1145/2883851.2883944.

[3] Chen, Q., Yue, X., Plantaz, X., Chen, Y., Shi, C., Pong, T.-C., and Qu, H., 2018. ViSeq: Visual Analytics of Learning Sequence in Massive Open Online Courses. IEEE transactions on visualization and computer graphics.

[4] Christiansen, R.B., Gynther, K., and Jørnø, R., 2019. The dinosaur that lost its head: A contribution to a framework using Learning Analytics in Learning Design. Tidsskriftet Læring og Medier (LOM) 12, 21, 21-21.

[5] Crosslin, M., 2018. Exploring self-regulated learning choices in a customisable learning pathway MOOC. Australasian Journal of Educational Technology 34, 1.

[6] Davis, D., Chen, G., Hauff, C., and Houben, G.-J., 2016. Gauging MOOC Learners' Adherence to the Designed Learning Path. In Proceedings of the International Educational Data Mining Society. (2016).

[7] Deboer, J., Stump, G., Breslow, L., and Seaton, D., 2013. Diversity in MOOC Students' Backgrounds and Behaviors in Relationship to Performance in 6.002x.

[8] Ferguson, R., 2012. Learning analytics: drivers, developments and challenges. International Journal of Technology Enhanced Learning 4, 5/6. DOI= http://dx.doi.org/10.1504/ijtel.2012.051816.

[9] Guàrdia, L., Maina, M., and Sangrà, A., 2013. MOOC design principles: A pedagogical approach from the learner's perspective. elearning papers, 33.

[10] Guo, P.J. and Reinecke, K., 2014. Demographic differences in how students navigate through MOOCs. In Proceedings of the first ACM conference on Learning@ scale conference, 21-30.

[11] Kivimäki, V., Pesonen, J., Romanoff, J., Remes, H., and Ihantola, P., 2019. Curricular Concept Maps as Structured Learning Diaries: Collecting Data on Self-Regulated Learning and Conceptual Thinking for Learning Analytics Applications. Journal of Learning Analytics 6, 3, 106-121.

[12] Kizilcec, R.F., Piech, C., and Schneider, E., Deconstructing disengagement: analyzing learner subpopulations in massive open online courses, 170-179.

[13] Koedinger, K.R., Mclaughlin, E.A., Jia, J.Z., and Bier, N.L., 2016. Is the doer effect a causal relationship?: how can we tell and why it's important. In Proceedings of the Sixth International Conference on Learning Analytics & Knowledge ACM, 388-397.

[14] Krathwohl, D.R. and Anderson, L.W., 2009. A taxonomy for learning, teaching, and assessing: A revision of Bloom's taxonomy of educational objectives. Longman.

[15] Lang, A., 2000. The limited capacity model of mediated message processing. Journal of communication 50, 1, 46-70.

[16] Law, N., Li, L., Herrera, L.F., Chan, A., and Pong, T.-C., 2017. A pattern language based learning design studio for an analytics informed inter-professional design community. Interaction Design and Architecture (s).

[17] Lei, S., Cristea, A.I., Awan, M.S.K., Hendrix, M., and Stewart, C., 2013. Towards understanding learning behavior patterns in social adaptive personalized e-learning systems AMCIS.

[18] Littlejohn, A., Hood, N., Milligan, C., and Mustain, P., 2016. Learning in MOOCs: Motivations and self-regulated learning in MOOCs. The Internet and Higher Education 29(2016/04/01/), 40-48. DOI= http://dx.doi.org/https://doi.org/10.1016/j.iheduc.2015.12.003.

[19] Liyanagunawardena, T.R., Adams, A.A., and Williams, S.A., 2013. MOOCs: A systematic study of the published literature 2008-2012. The International Review of Research in Open and Distributed Learning 14, 3, 202-227.

[20] Lockyer, L., Heathcote, E., and Dawson, S., 2013. Informing Pedagogical Action: Aligning Learning Analytics With Learning Design. American Behavioral Scientist 57, 10, 1439-1459. DOI= http://dx.doi.org/10.1177/0002764213479367.

[21] Maiyuran, J., Bajwa, A., Bell, A., Hemberg, E., and O'reilly, U., 2019. How Student Background and Topic Impact the Doer Effect in Computational Thinking MOOCs. In 2019 IEEE Learning With MOOCS (LWMOOCS), 47-52. DOI= http://dx.doi.org/10.1109/LWMOOCS47620.2019.8939643.

[22] Marzouk, Z., Rakovic, M., Liaqat, A., Vytasek, J., Samadi, D., Stewart-Alonso, J., Ram, I., Woloshen, S., Winne, P.H., and Nesbit, J.C., 2016. What if learning analytics were based on learning science? Australasian Journal of Educational Technology 32, 6.

[23] Mor, Y., Ferguson, R., and Wasson, B., 2015. Editorial: Learning design, teacher inquiry into student learning and learning analytics: A call for action. British Journal of Educational Technology 46, 2, 221-229. DOI= http://dx.doi.org/10.1111/bjet.12273.

[24] Papamitsiou, Z. and Economides, A.A., 2014. Learning analytics and educational data mining in practice: A systematic literature review of empirical evidence. Journal of Educational Technology & Society 17, 4, 49-64.

[25] Persico, D. and Pozzi, F., 2015. Informing learning design with learning analytics to improve teacher inquiry. British Journal of Educational Technology 46, 2, 230-248. DOI= http://dx.doi.org/10.1111/bjet.12207.

[26] Shi, C., Fu, S., Chen, Q., and Qu, H., 2014. VisMOOC: Visualizing video clickstream data from massive open online courses. In 2014 IEEE Conference on Visual Analytics Science and Technology (VAST) IEEE, 277-278.

[27] Shibani, A., Knight, S., and Shum, S.B., 2019. Contextualizable Learning Analytics Design. In Proceedings of the Proceedings of the 9th International Conference on Learning Analytics & Knowledge - LAK19 (2019), 210-219. DOI= http://dx.doi.org/10.1145/3303772.3303785.

[28] Siemens, G., 2013. Learning Analytics: The Emergence of a Discipline. American Behavioral Scientist 57, 10, 1380-1400. DOI= http://dx.doi.org/10.1177/0002764213498851.

[29] Siemens, G. and Long, P., 2011. Penetrating the fog: Analytics in learning and education. EDUCAUSE review 46, 5, 30.

[30] Simon, M.A., 1993. Reconstructing mathematics pedagogy from a constructivist perspective.

[31] Sinha, T., Jermann, P., Li, N., and Dillenbourg, P., 2014. Your click decides your fate: Inferring information processing and attrition behavior from mooc video clickstream interactions. arXiv preprint arXiv:1407.7131.

[32] Stöhr, C., Stathakarou, N., Mueller, F., Nifakos, S., and Mcgrath, C., 2019. Videos as learning objects in MOOCs: A study of specialist and non-specialist participants' video activity in MOOCs. British Journal of Educational Technology 50, 1 (2019/01/01), 166-176. DOI= http://dx.doi.org/10.1111/bjet.12623.

[33] Sztajn, P., Confrey, J., Wilson, P.H., and Edgington, C., 2012. Learning Trajectory Based Instruction. Educational Researcher 41, 5, 147-156. DOI= http://dx.doi.org/10.3102/0013189x12442801.

[34] Wang, Y., Law, N., Hemberg, E., and O'reilly, U.-M., 2019. Using detailed access trajectories for learning behavior analysis. In Proceedings of the 9th International Conference on Learning Analytics & Knowledge, 290-299.

[35] Wise, A.F., Vytasek, J.M., Hausknecht, S., and Zhao, Y., 2016. Developing Learning Analytics Design Knowledge in the" Middle Space": The Student Tuning Model and Align Design Framework for Learning Analytics Use. Online Learning 20, 2, 155-182.