Toward Large-Scale Learning Design

Categorizing Course Designs in Service of Supporting Learning Outcomes

Dan Davis*
Delft University of Technology
Delft, the Netherlands
d.j.davis@tudelft.nl

Daniel Seaton
Harvard University
Cambridge, MA, USA
daniel_seaton@harvard.edu

Claudia Hauff
Delft University of Technology
Delft, the Netherlands
c.hauff@tudelft.nl

Geert-Jan Houben
Delft University of Technology
Delft, the Netherlands
g.j.p.m.houben@tudelft.nl

ABSTRACT

This paper applies theory and methodology from the learning design literature to large-scale learning environments through quantitative modeling of the structure and design of Massive Open Online Courses. For two institutions of higher education, we automate the task of encoding pedagogy and learning design principles for 177 courses (which accounted for for nearly 4 million enrollments). Course materials from these MOOCs are parsed and abstracted into sequences of components, such as videos and problems. Our key contributions are (i) describing the parsing and abstraction of courses for quantitative analyses, (ii) the automated categorization of similar course designs, and (iii) the identification of key structural components that show relationships between categories and learning design principles. We employ two methods to categorize similar course designs—one aimed at clustering courses using transition probabilities and another using trajectory mining. We then proceed with an exploratory analysis of relationships between our categorization and learning outcomes.

INTRODUCTION

The ubiquity of digital learning platforms is leading to new ways of documenting and understanding course design. Even though online learning platforms often constrain instructors to design choices in the limited context of videos, text, and various assessment components, there still exists a vast and uncharted diversity in the way instructors choose to design and structure their digital learning materials. A recent example in scaled learning is the edX consortium, where over 1,700 courses[1] have been created by 118 institutions[2] across the globe. This makes for a truly massive possibility space that spans discipline, culture, and pedagogy.

MOOC researchers have begun analyzing course design and pedagogy in order to understand this diversity, but the work has been isolated and largely a process of human categorization based on broad interpretations of learning design. Recent applications of pedagogical inventories involving human classification on a number of scales exemplifies these efforts. The authors in [24] have compared the pedagogical structure of 17 MOOCs using an inventory called AMP (Assessing MOOC Pedagogies), and [17] applied a similar inventory across 78 MOOCs. Both found signs that many MOOCs are replicating traditional instruction tactics. Such work can potentially help address best practices in course design, but it has remained a manual task and not yet found widespread adoption by researchers in the MOOC community.

Furthermore, researchers in more traditional areas of learning design have only been able to conduct small-scale (usually on a single, course-by-course basis) mostly-qualitative analyses of course structures and their relationship with learning outcomes. And although the number of courses considered is small (typically ranging from 1–20 [6]), learning designers have developed methods for comparing and classifying courses' structures. This is achieved through a process of *abstraction*, or the separation of a course's topical content (such as math, engineering, history, etc.) and its internal structure (the sequence of activities used to teach the content).

In another area of research, learner behavior modeling has taken off for MOOCs [4, 7, 8, 25]. However, there has yet to be any large-scale or automated evaluation of the effectiveness of various learning design patterns using the tools from the learner-behavior community. So while there is a quickly emerging corpus of learner modeling research unfolding, there have not yet been any empirical efforts to connect the findings to the design of the learning environment.

*The author's research is supported by the *Leiden-Delft-Erasmus Centre for Education and Learning.*

L@S'18, June 26–28, 2018, London, England, UK

ISBN 123-4567-24-567/08/06...$15.00
DOI: 10.1145/1235

[1] https://www.edx.org/course?course=all
[2] https://www.edx.org/schools-partners

By understanding the theory and methodology from the learning design literature and applying it to large-scale learning environments, we are able to advance the field of learning at scale through a quantitative analysis of the structure/design of online learning environments. The MOOC community has primarily focused on learners so far. But the number of courses accessible to researchers is growing large enough to offer a new paradigm for teaching research at scale.

In this paper, we attempt to build a framework that can help aid classification of course design in an automated and scalable fashion. Our framework is largely built around the following ideas:

- A methodology to parse and abstract a course to enable quantitative analyses of its structure.
- Quantitative measurement of the difference between course designs.
- Identification of key structural components that differentiate courses with clustering and then gaining a deeper understanding through qualitative analysis.

Using a dataset made up of 177 MOOCs from two institutions of higher education, we abstract course design into a sequence of learner activities and apply two types of pattern mining, namely, (i) transition probability mining and (ii) trajectory mining. We explore both methods on an institution by institution basis. In addition, we explore the relationship between our classification (clusters) with a straightforward learning outcome — verified learner pass rates. This exploratory addition to the study is to further support whether our abstraction and automation can lend itself to goals of improving learning outcomes through better design.

RELATED WORK

Below we describe the current state of the art in the domains of learning design and learner behavior. Our review of the literature finds a distinct common thread connecting learning design and learner behavior studies, namely, that of abstraction and complexity reduction. In addition, many of the methods in our work are inspired by research in the area of learner-behavior pattern mining [4, 7, 8, 25]; we find that many methodologies in this field have have potential applications to pattern mining of course structure and pedagogy.

Learning Design Patterns

The learning design literature offers a substantial body of research theorizing about the design of learning patterns and sequences. Reference [11] offers an exhaustive review of how Learning Designers have tackled the challenge of describing and synthesizing *patterns for learning*, defined by [18] as a semi-structured description of a strategy for teaching a given topic or skill. The primary purpose of patterns for learning is to externalize knowledge in a way that can be generalized and accessed by members of the teaching community.

Traditional classroom teaching environments do not require explicit documentation of a strategy or pattern for teaching. These are often proprietary and documentation standards vary across institutions [5, 11]. Reference [11] calls for a standardization to facilitate sharing of patterns for learning throughout the teaching community in having teachers “enact design science” as a normal part of the teaching practice so that, as a community, they can gain an understanding over which designs lead to which outcomes/achievements.

One effort to facilitate the comparison and standardization of teaching design patterns is found in [5] where the authors developed a “Teaching Method Template” which describes instruction primarily in terms of activity sequences—found to be the most effective method of depicting patterns for learning in terms of teacher preference and usefulness.

In this template, reference [5] represents activity sequences both textually and graphically: the *graphical representation* uses flow charts and activity diagrams to visualize patterns in a way users can quickly internalize and the text-based *sequence of activities* approach details the temporal sequence of activities and assessments in a given plan. The authors in reference [11] identify the “sequence of activities” approach (defined as a collection of teaching design patterns building towards an outcome) as the most interesting and promising in the age of digital learning and instruction. This includes the decisions of which activities to introduce at which point, but also the effective transition between activities so that each activity appropriately informs the next. Though this topic is not yet prevalent in the area of digital learning environments, the authors in [11] claim that “The origin or provenance of a pedagogical pattern is as important as citations are in research. Teachers considering adopting a new pattern need to know its origin, and should be able to track the way it has developed into alternative versions.” There is not yet a widely accepted standard for patterns as of yet, however, digital learning environments present a tremendous opportunity to develop, track, and share pedagogical patterns due to how content is stored in digital-learning platforms, as was done in [20].

Teaching design patterns that have been identified as topic-specific are referred to as “signature pedagogies”[23]. An example of signature pedagogies is the contrast between hands-on (bedside) teaching for medical education and the inquisitive nature of a law school lecturer (firing off strings of question sequences to their audience members). To the best of our knowledge, no work as of yet has been done to evaluate the detailed patterns of such signature pedagogies in the context of digital-learning environments.

Reference [12] poses this question about the extent to which disciplines can be “disentangled” from their signature pedagogies. This leaves the question open about whether some strategies are best kept tied to a specific discipline, or perhaps through the sharing of such pedagogical wisdom, disciplines can benefit from a new perspective. [11] introduces a method of documenting instructional sequences in a structured and standardized manner so that patterns from one domain “can be replaced with entirely new topic content to generate the same pattern in a different subject area.”

The key to this “disentangling” of pedagogies from their disciplines is a successful abstraction of the pedagogy to a form that is transferable to a new context. And by removing the content and only focusing on the activity type and transitions between activities, we arrive at a structured method of documenting patterns and sequences for learning [10, 13].

Learner Behavior Patterns

We next describe methods from research in learner behavior patterns and their applications to the above challenges of learning design patterns. There has recently been a surge in research exploring MOOC learners' navigational patterns throughout course activities. The impact of this research stems from our ability to see learner behavior in highly self-directed environments, i.e., without instructor oversight. However, while these methods continue to be evaluated and developed in the context of learner activity patterns and navigational events, we here propose that similar methods ought to be employed in evaluating course design patterns in digital learning environments. Doing so will allow us to better understand how course design and the sequencing of activities are related to learner behavior. Below we review work on learner behavior modelling while pointing out how previous work influences our methods for course structure pattern mining.

The research presented in [4, 7, 8, 25] characterizes MOOC learners through their clickstream data tracking their transition between activities. Reference [25] first identified common 2-gram event transitions; [4] next extended these to 8-gram event sequences and labeled the sequences as various motifs representing a study pattern; and [8] extended this by connecting these event transitions to self-regulated learning strategies using learner self-reported survey results as well. [7] builds on the Markov modeling technique in [4] by developing a two-layer Markov model which accounts for transitions between both micro and macro activity patterns. In the present research we apply this methodology of analyzing learner transition probabilities to course structure data—exploring the transitions between course components as defined by the instructor as opposed to the path executed by the learner.

Reference [2] builds upon the work in [4, 8, 25] by applying clustering techniques to MOOC learner behavior. Clustering in this case enables the automatic identification of similar trajectories to be identified at scale, whereas prior work in this area was done manually [17, 24]. We apply this scalable clustering approach to MOOC course structures in the present research. Reference [2] employed both pattern- and data-driven approaches for analyzing and clustering MOOC learner activity data. They correlated learner engagement patterns with course learning outcomes as well—final course grades earned and each cluster's overall passing rate. The authors first categorize learners into one of four categories (separated by behavior patterns preceding assessment) on a week-by-week basis to account for changes over time, and then they use hierarchical agglomerative clustering to group learners with similar week-by-week trajectories.

The authors in [2] also introduce a second method to track latent learner activity patterns with an unsupervised processing pipeline. The pipeline is comprised of four phases: (i) activity sequence modeling, where a transition matrix is generated and used as a learner model, (ii) distance computation, (iii) clustering, where the dissimilarity matrix is clustered with hierarchical agglomerative clustering using the Ward's method, and (iv) cluster matching, to identify temporal relationships between identified clusters. Based on this method, the authors enable a direct comparison of various types/patterns (clusters) of behavior and academic achievement, very similar to our method presented here of clustering course structures.

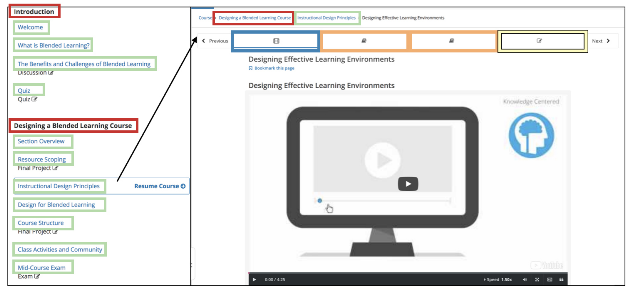

Figure 1. edX platform screenshot with components and containers associated with the OLX format marked by color: chapter (red), sequential (green), videos (blue), html (orange), and problems (yellow).

The primary methodology employed in [16, 19] is that of process mining. Such process mining techniques include (i) visualization, where processes are plotted in a variety of graph types in order to make trends and patterns visually apparent, (ii) conformance checking, where actual/executed processes are compared to the normative/intended model, and (iii) process discovery, where a process model is learned from event log data. Whereas the authors in [2] were motivated by connecting behavior to learning outcomes, the authors in [16, 19] are motivated to model learner behavior in order to develop targeted interventions to support learners in developing self-regulated learning skills.

After reviewing the state of the art and existing knowledge gaps in the literature of learning design patterns and learner behavior patterns, we arrive at the following three primary **R**esearch **Q**uestions:

RQ1 To what extent can we parse and abstract the design of a MOOC by employing principles from the learning design literature?

RQ2 How can we quantitatively compare and contrast the design of MOOCs?

RQ3 Are there structural components that differentiate a MOOC's design?

In addition, we put forward an exploratory **RQ4** addressing the relationship between our abstraction of course design and students' learning outcomes.

METHODS

Building upon the learning design methodology of abstracting a course's structure away from content, the present research methodology employs an exploratory approach in applying methods from research in learning design and learner behavior patterns to the topic of learning design patterns in digital learning environments. We next outline the methodology used with regard to each guiding research question.

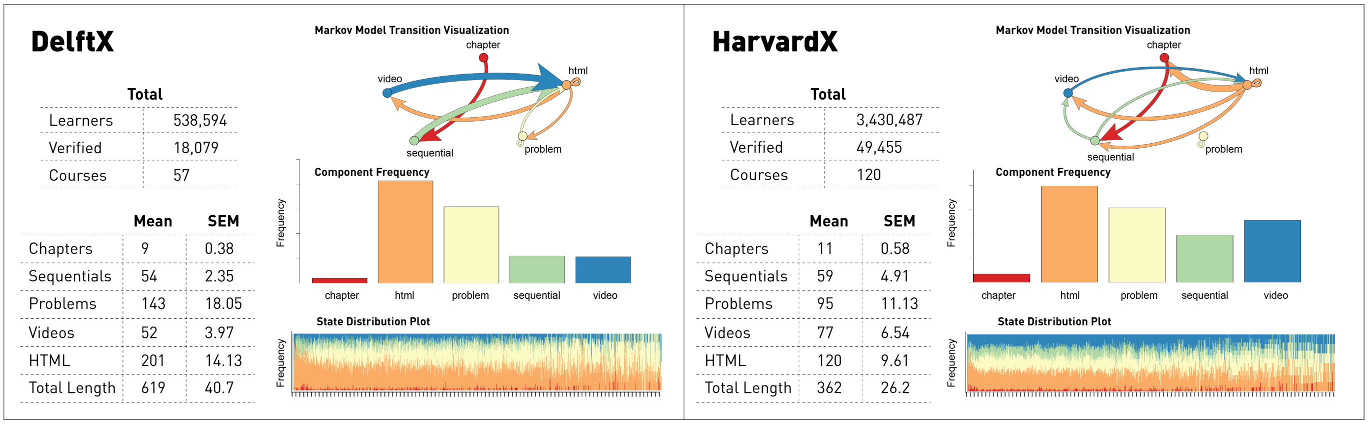

Figure 2. Course structure overview for each institution. Tables indicate the total number of enrolled and verified learners for each institution, along with summary statistics about the occurrence of course components (mean per course and standard error of the mean (SEM)). The Markov model transition visualization indicates the most common event type transitions across all courses for each institution; edge/line weights distinguish transition prominence. Component frequency bar graphs show how common each component type was across all courses. The state distribution plot – depicting the left to right occurrence of course components – is a trajectory mining visualization that accounts for the likelihood of component occurrence accounting for all courses in each institution.

Dataset
Our dataset consists of edX MOOCs from Delft University of Technology (or DelftX, as it is known on the edX platform) and Harvard University (HarvardX). Within this study, DelftX accounts for 57 MOOCs with a total of 35,283 course components, and HarvardX accounts for 120 MOOCs with a total of 43,514 components. In edX courses content can be broken up into components and collections. Components are stand-alone assets with which learners interact: videos, problems, html pages, and custom activities. Collections are containers that provide structure and navigation for learners: chapters, sequentials, and verticals. For clarity, all components and collections are illustrated in context in Figure 1.

In this study, we remove verticals from consideration to reduce complexity, namely, in our own ability to interpret results, as the institutions studied tend to have short verticals, leading to numerous verticals that act as delimiters between small numbers of resources. While future analyses can include verticals, we found here that verticals in DelftX courses typically include between 2 and 3 resources (avg. of 2.75 resources per vertical) and, for HarvardX courses, 1 to 3 resources per vertical (avg. 1.7 resources per vertical). We omit verticals to allow for an analysis of longer, more representative learning design sequences. We also omit custom components, which have extreme variation in students' interactions and in many cases evolve over time (i.e., may not have the same use case from course to course). In addition, the namespaces for these components do not remain consistent, making them difficult to track in this initial study.

Parsing edX Courses
All content authored for the edX platform is stored in the Open Learning XML (OLX) format[3]. OLX is a standard that allows the transfer of content between instances of the open source edX platform, authorship outside the platform, and extraction of information related to course design (like in this work). OLX contains the raw markdown (XML) for all authored content in a course, namely, all content tags, text associated with content, and relevant metadata. Courses are generally designed in edX Studio – a GUI for creating and structuring courses – masking the OLX from most users. OLX data can be exported through edX Studio and is also provided in regular data exports to edX consortium members through the edX research pipeline [4]. For each course in the present study, we download the OLX data and pass it through a parsing algorithm to structure the data in a more desirable format for analysis (colloquially referred to as the "course axis"). All OLX components are sorted in sequential order according to their placement in the course.

Abstracting Structure from Content
Research in learning design relies heavily on the process of abstracting a course structure into a standardized, comparable structure. Abstraction here is the process of stripping away the course topic materials from the underlying structure and components (**RQ1**). For example, in a course about Statistics, a given sequence of activities might include: a lecture about the difference between frequentist and bayesian statistics → discussion about the benefits and drawbacks of each approach → exam assessing learners' ability to apply what they've learned. The abstracted version of this sequnce would become: lecture → discussion → assessment. This method for abstraction is also commonly used when considering learner activity in courses as well [2, 4, 21, 25]. We view this abstraction as similar to processes like coarse-graining in physics, where microscopic structure is often approximated in order to measure macroscopic properties of a system.

Computing Similarity
After abstraction of a course, we qualitatively measure the differences between course structures (**RQ2**) using two approaches: (i) clustering transition probability, and (ii) trajectory mining. Transition probability treats the course activity

[3] http://edx.readthedocs.io/projects/edx-open-learning-xml

[4] https://github.com/edx/edx-analytics-pipeline

1. Original Sequence

A B C A C B A B

2. Edge List

A B C A C B A B

A B C A C B A B

3. Transition Matrix (P)

	A	B	C
A	0.0	0.66	0.33
B	0.5	0.0	0.5
C	0.5	0.5	0.0

4. Sample Transition Matrix (Q)

	A	B	C
A	0.25	0.5	0.25
B	0.0	0.7	0.3
C	0.5	0.5	0.0

5. Distance Matrix (| P - Q |)

	A	B	C
A	0.25	0.16	0.08
B	0.5	0.7	0.2
C	0.0	0.0	0.0

Figure 3. The process of calculating similarity using transition probability. (1) Original sequence of elements, read from left to right. (2) Edge list: showing the transitions from each source element to its target. (3) Transition matrix showing the probability of each source element's (y-axis) transition to each target element (x-axis). Each row sums to 1.0. (4) Sample transition matrix (Q) to compare to P. (5) Distance matrix for P and Q.

sequence as a Markov chain and considers the prominence of each of the possible transitions between activity types. The choice for this approach is based on the learning design principle which highlights the importance of the *consecutive* sequencing of learning activities. The trajectory mining approach takes the entire sequence into account by calculating differences in the order and position of all components, which allows for the analysis of learning design sequences over the span of entire courses beyond single transitions.

We employ both methods for computing dissimilarity between course structures because both have been used in prior research for learning path analysis [4, 8, 16], and both methods have their own advantages and drawbacks. For example, the main advantage transition probability has over trajectory mining is that the length of the sequence is not considered , whereas in trajectory mining the difference in sequence length imposes a significant bias/cost on the results. On the other hand, the main benefit trajectory mining has over transition probability is that it takes the entire course sequence into consideration and enables more macro-level course design insights.

Transition Probability

A transition matrix is a method of representing a sequence of transitions, or a Markov chain. Computing a transition matrix has been a prominent method for modeling learner behavior in online learning environments [4, 8, 16], but this method has not yet been applied to teaching or instructional behavior. By adopting a method focusing on transitions from one activity to the next, we are able to connect digital learning environments to the literature on learning design.

We compute transition matrices by first generating an edge list, as shown in Figure 3.2 which represents all origin→target pairings (sequential connections from one event type to the next) from the original sequence of elements from Figure 3.1. This edge list is then used to compute the probability of each event type transitioning to the next, and these proportions are then used to populate the final transition matrix.

We generated transition matrices (P and Q) for all 177 courses included in the study and stored them in a list of matrices. For each institution, we generate transition matrices for each course. We then calculate the L_1 distance (also referred to as Manhattan distance or taxicab metric) (d_1) between transition matrices ($P - Q$) on a course by course basis and sum the absolute values between them:

$$d_1(P,Q) = ||P-Q||_1 = \sum_{i=1}^{n} |P_i - Q_i|$$

1. Original Sequences

1. A B C D E
2. A C C D
3. A B C C C

2. OM Distances

	1	2	3
1	0.0	3.0	4.0
2	--	0.0	5.0
3	--	--	0.0

Figure 4. The process of calculating OM distance. (1) Original three sequences of elements, read from left to right. (2) Matrix showing the OM distance between sequences as the cheapest total editing cost.

where P and Q are transition matrices flattened into one-dimensional vectors.

For example, the distance between P and Q, $d_1(P,Q)$, in Figure 3 amounts to 1.89. The final distance matrix contains each of these calculated differences for all matrix pairings and is then is a suitable format to be clustered—noting that all matrices must contain the same columns and rows to ensure appropriate calculations.

Trajectory Mining

The trajectory mining method first computes a distance matrix using the optimal matching (OM) method. This distance matrix is populated by edit distances (or the minimal editing cost): the minimal cost of all insertions, substitutions, and deletions to transform one sequence into another [14]. In accordance with the method introduced in [14], substitutions (C_S) have an editing cost of 2.0 and insertions & deletions (C_I) have an editing cost of 1.0. The editing costs according to [14] are:

$$C_S = 2 - p(i|j) - p(j|i) \quad \text{and} \quad C_I = 1 - p(i|j) - p(j|i)$$

where $p(i \mid j)$ is the transition rate between states i and j.

Figure 4 illustrates the process of arriving at the distance matrix between two sequences with a substitution colored in blue and an insertion colored in orange for sequences 1 and 2.

Clustering Similar Courses

In service of **RQ3**, we uncover similarities in courses' structures by employing Ward's method of hierarchical agglomerative clustering. This method starts by considering all courses as n independent clusters. The algorithm progresses by forming $n-1$ clusters and computing the error sum of squares and r^2 value at each step. Clusters are then formed by grouping units which yield the lowest sum of squares and highest r^2 values. When all n units are combined into a single large cluster tree (or dendrogram), the algorithm stops.

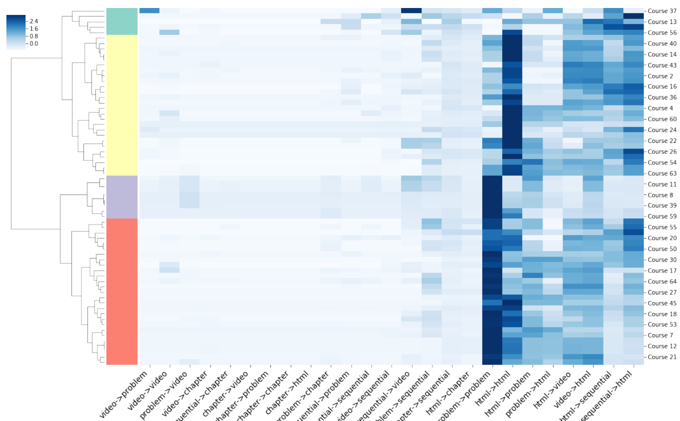

Figure 5. Hierarchically clustered heatmaps using transition probabilities as input for DelftX. The color of each cell shows the prominence of a transition (x-axis; darker colors are higher values). To the left is a dendrogram showing the agglomerative clustering process, and the leftmost column shows which cluster each course (y-axis) belongs to. Course titles withheld for blind review. Best viewed in color.

Once we have completed Ward's hierarchical clustering method, we then determine the optimal number of clusters within that single tree. To do so we employ the Calinski-Harabasz index method [3]. This method evaluates the validity of clusters according to the average within-cluster sum of squares and the average between-cluster sum of squares [15]. The index aims to maximize both the distance between cluster centers as well as the individual cluster compactness. We next verified this result by calculating the silhouette scores of each clustering result, an alternative method for measuring cluster tightness and separation [22].

These approaches are a common and widely accepted way of uncovering trends in large datasets [1, 3, 15] and have been successfully applied to large-scale learning problems in the past [2]. Based on the results of these analyses, we then address **RQ3** by drawing semantic meaning through qualitative analyses of the clusters.

Exploring Course Learning Outcomes

After developing an understanding of common course designs, we next explore the extent to which similar course designs are related to learning outcomes (**RQ4**). To evaluate in an exploratory fashion whether there are statistically significant differences in completion rates between clusters, we fit a one-way ANOVA model considering course completion rates among verified learners (those who went through a process to verify their identity with edX) by cluster group.

RESULTS

Abstracting Structure from Content

In service of **RQ1**, we find that our abstraction of courses sufficiently enables qualitative insights into course design decisions. For example, the authors on this paper from HarvardX can confirm an abundance of video in Figure 2, as reflected in the bar graph. Additionally, HarvardX pivoted toward smaller, modular courses. In some cases, taking long 16 week courses and breaking them up into multiple course—reflected in the average course length. For DelftX, which offers predominantly STEM courses, we confirm a trend towards longer courses containing more assessment activities.

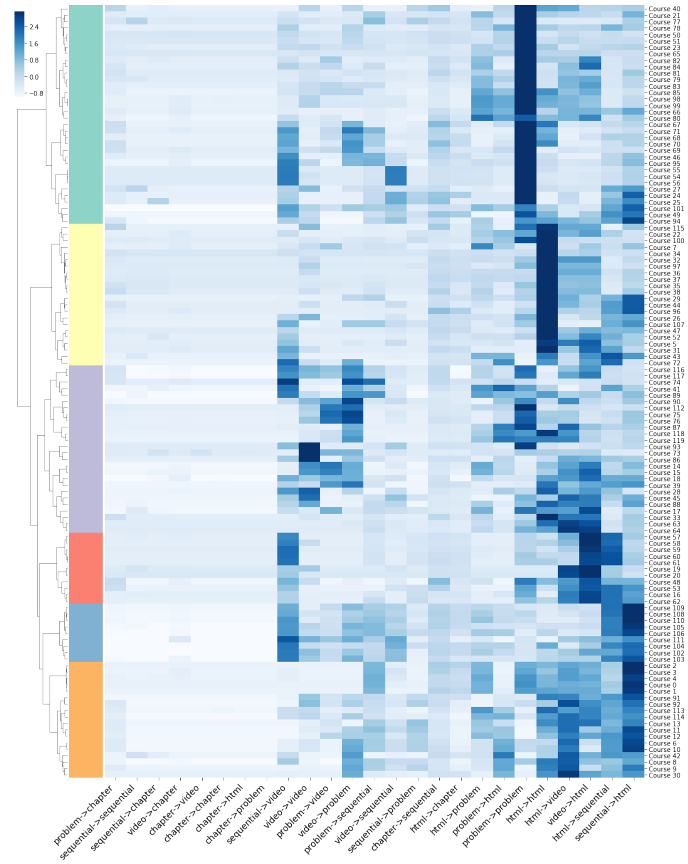

Figure 6. Hierarchically clustered heatmaps using transition probabilities as input for HarvardX.

In the following analyses, we draw the following connections between the syntactic form of the OLX format and the semantics of learning design: *chapter* and *sequential* elements indicate a section break in the course continuity. Sequentials house subtopics of chapters and are used to break up material into manageable chunks for learners. *Video* components are indicative of video lecture activities and are the primary method for introducing learners to new content or concepts. *Problem* elements are used as graded assessment events where learners are given the opportunity to test their newly gained knowledge. Lastly, *html* elements are used to help guide the learner between video lectures and assessments and provide navigational guidance/context.

From this method, we find evidence that despite the limited number of elements available in an online learning platform like edX, substantial variation does indeed occur in the learning and structural design of various courses.

Clustering Similar Course Structures

The following results address the quantitative comparison of course structures toward **RQ2**. Figures 5 and 6 visualize the

Table 1. The percentage of transition types for all courses within clusters for both institutions. The bottom row indicates the total number of courses included in each cluster. Only the most prominent transition types /factors are shown.

		DelftX				HarvardX					
		Clust. 1	Clust. 2	Clust. 3	Clust 4	Clust. 1	Clust. 2	Clust. 3	Clust. 4	Clust. 5	Clust. 6
% All Transitions	html→html	10.5	18.9	8.7	8.5	5.8	31.7	8.8	2.9	3.7	10.2
	html→problem	0.0	4.8	1.7	2.4	6.0	4.9	4.9	0.2	4.8	5.9
	html→video	10.5	15.4	11.2	14.9	5.2	8.1	9.6	17.4	3.7	13.9
	problem→html	2.6	9.7	6.9	10.7	5.1	4.4	5.3	0.4	3.8	5.4
	problem→problem	1.3	12.8	61.1	34.6	32.8	7.7	12.2	7.0	3.0	4.2
	sequential→html	31.6	14.1	2.9	7.6	7.3	7.7	5.4	9.2	15.0	15.5
	sequential→problem	7.9	4.0	2.5	4.1	1.5	0.8	0.9	2.6	2.5	2.9
	sequential→video	5.3	0.4	0.0	0.2	6.0	4.0	5.0	9.2	10.8	1.3
	video→html	10.5	12.8	0.4	12.2	5.1	10.2	9.1	26.1	4.8	9.7
	video→video	1.3	0.4	0.0	0.2	2.6	2.4	7.8	1.4	4.5	2.0
	# Courses	4	23	6	24	34	21	25	11	9	18

transition probability features (color-map, where darker cells are larger values) and the dendrogram based on our agglomerative clustering approach. Clusters are indicated by color in the leftmost column of each figure, namely, 4 clusters for DelftX in Fig. 5 and 6 clusters for HarvardX in Fig. 6.

In identifying the ideal cluster number for the transition probability method we relied on the Calinski-Harabasz index [3] and silhouette [22] method, along with viewing our dendrograms (y-axis of Figures 5 and 6) for sensible cutoffs [9].

To determine the optimal number of clusters to use with the trajectory mining approach, we again computed clustering quality measures using the Calinski-Harabasz index [3] and silhouette [22] method. We determined the optimal number of clusters for DelftX to be four and for HarvardX to be three.

Key Structural Components

With regard to **RQ3** which is concerned with identifying the key structural components that define each cluster of similar courses based on quantitative analyses of their syntactic structure, we highlight the qualitative insights offered by each method into the semantic trends which define each cluster. By contextualizing each element into its place in the course relative to other elements, we identify learning design patterns that distinguish each category.

Transition Probability: DelftX

With regard to DelftX, Figure 5 shows two key transition types with prominent transition rates correlated to clusters, namely, *problem-problem* and *html-html*, both indicated by darker color. These are in contrast to the less-prominent transitions found in the left portion of the graph (such as *video-problem* transitions). The cluster map indicates that some transition rates have larger effects than others.

The most dominant feature in **Cluster 1** (green) is the *sequential-html* transition type which accounts for 31.6% of all transitions in the cluster, indicating frequent use of text/reading activities to introduce new sequences. Another prominent feature of is the proportion of *video-video* events, which account for 1.3% of all transitions. And even though this indicates a low prominence of consecutive video lectures, it is the highest among clusters from University A (but lower than any HarvardX cluster; to be discussed).

In **Cluster 2** (yellow), the *html-html* transition type accounts for 18.9% of all transitions, indicating a substantial amount of consecutive reading activities. And with *html-video* also being a dominant feature in this cluster, we see that those sequences of consecutive reading activities are often followed by a video lecture activity. Also worth noting is the trend that any transition involving html elements/reading activities is high in this cluster, indicating that, regardless of context, courses here are comprised mainly of reading activities.

The *problem-problem* transition type is the most prominent feature of **Cluster 3** (purple) in accounting for 61.1% of all transitions in the courses. That is nearly twice as prominent as any other transition frequency from either institution. We may assume long assessment activities to be the main function of the courses in this cluster. There are no chains of consecutive video lectures, and sections never begin with videos.

The *problem-problem* transition type is also a dominant feature in **Cluster 4** (red), but this cluster is distinguished from Cluster 3 with its relatively high frequency of *video-html* transitions. While Cluster 3 contained very few video lecture activities, Cluster 4 strikes a closer balance of being assessment heavy while still offering more video lecture activities. From this transition we further note that reading activities typically follow video lectures, likely providing a summary or preparing learners for the next assessment activity.

Transition Probability: HarvardX

With regard to HarvardX, Figure 6 shows more clusters and more variation among clusters. While containing a largely even distribution of most transition types, **Cluster 1** (green) is dominated by the *problem-problem* feature, which accounts for 32.8% of all transitions in this cluster. This indicates that, similar to Clusters 3 and 4 in DelftX, these courses contain numerous long assessment activities. Another trait of courses in this cluster is the relatively high prominence of the *html-problem* feature. This indicates that courses in this cluster most often preface their assessment activities with a reading activity.

In **Cluster 2** (yellow), 31.7% of all transitions among courses are *html-html*, meaning that courses in this cluster have frequent, extended strings of consecutive reading activities. Also prominent at 10.2% of transitions is the *video-html* type. From this we infer that those long strings of reading activities are often preceded by lecture video activities.

The most distinguishing feature of **Cluster 3** (purple) is *video-video* transition types, accounting for 7.8% of all transitions in this cluster. Across institutions, this cluster has the highest frequency of consecutive video activities. Video lectures are the primary method of instruction here, and we also see these lectures are typically followed by strings of consecutive assessment activities (accounting for 12.2% of all transitions in the cluster).

Cluster 4 (red) is primarily characterized by prominent *video-html* transition types (26.1%), which is more than twice the frequency of any other cluster for this transition type. This indicates that video lecture activities for courses in this cluster are most frequently followed by reading activities. And given that the most common transition from reading activities is to video lecture activities (*html-video* 17.4%), we can see that courses in this cluster often adopt the pattern of alternating video lecture and reading activities.

The most unique trait of **Cluster 5** (blue) is a relatively even distribution of all event transition types. The two most prominent are *sequential-html* and *sequential-video*, at 15.0% and 10.8% respectively. This may indicate that new sections in these courses are typically introduced with either reading or video lecture activities, noting the frequency of new sections beginning with video activities in this course is the highest among clusters from HarvardX.

The *sequential-html* (15.5%) transition type is also the most prominent in **Cluster 6** (orange), but this cluster is differentiated from Cluster 5 by its low *sequential-video* transition type (1.3%), which this is the lowest of any cluster from this institution. Also prominent in this cluster is the high frequency of the *html-video* transition type, which shows a mixing of video lectures and reading activities.

Trajectory Mining: DelftX

In Figures 7 and 8 we observe that the clustering results from the trajectory mining approach are primarily influenced by (i) the length of a given sequence, (ii) the frequency of activity types within a sequence, and (iii) the temporal order/placement of activities within the entire course trajectory, whereas the clustering based on transition probabilities (above) was illustrative of the sequence and order of activity types.

With regard to DelftX (Figure 7), in **Cluster 1** we observe mostly reading (representing 42% of all activities) and assessment activities (34%). One interesting characteristic of this cluster's learning design is the equal frequency of section breaks and video lecture activities (each at 11%). This indicates that each section in the course consists of a single video lecture activity. Reading activities are most prominent at the beginning of courses in this cluster.

The most distinguishing trait of **Cluster 2** is the prominence of assessment activities (48%) found throughout the sequence, with some notable large spikes in frequency throughout the courses—indicative of long exams and problem sets. Reading activities are also prominent at the beginning of this cluster

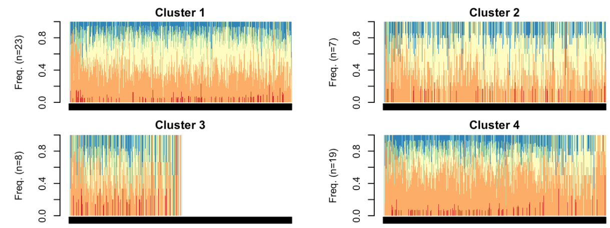

Figure 7. State distribution plots from the trajectory mining method per cluster for DelftX showing, at each point along the x-axis, the proportion of each component type.

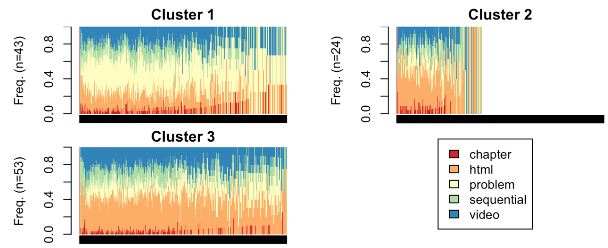

Figure 8. State distribution plots per cluster for HarvardX.

of courses. With the ratio of video lectures (12%) to section breaks (6%) strongly favoring the former, we observe that, unlike Cluster 1, each section is typically made up of two video lecture activities.

Cluster 3 is clearly characterized primarily by its short length, being on average half the length of others. The cluster is also quite noisy—lacking any discernable patterns. While the state distribution plot may not be the most illustrative due to its length, we do observe activity frequencies largely comprised of reading activities (42%).

The state distribution for **Cluster 4** indicates reading activities to be the prominent activity (54%). There are three times as many reading activities as there are assessments (18%) and very few video lecture activities (11%). We also observe high frequency of reading activities at the beginning of these courses, which indicates design patterns where introductory texts are used to prime learners.

Trajectory Mining: HarvardX

With regard to HarvardX (Figure 8), in **Cluster 1** we observe mostly assessment activities (40%) followed by a relatively high frequency of video lecture activities (20%). The general trajectory of these courses can be understood as designs of short introductory reading activities at the beginning of the course followed by long sequences heavy with assessment activities with the sporadic video lecture mixed in.

Similar to DelftX's Cluster 3, **Cluster 2** consists of courses with short length (on average 250 components). In addition, it appears noisy and without clear patterns from the visualization. However we observe that it is largely comprised of reading activities (39%) with very few video lectures (22%).

Cluster 3 contains courses with a high frequency of video lectures (23%) and reading activities (45%). As is the case with Cluster 2 from DelftX, these are the only clusters with

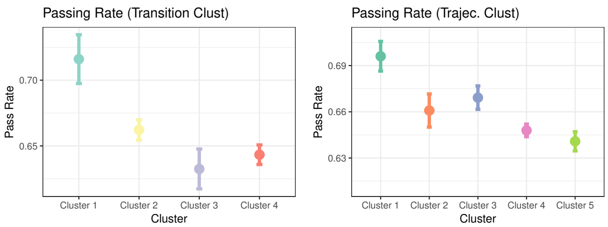

Figure 9. The mean and SEM (error bars) of passing rates of each cluster from DelftX courses.

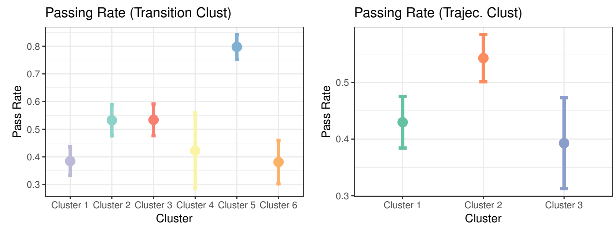

Figure 10. The mean and SEM (error bars) of passing rates of each cluster from HarvardX courses.

more videos than problems, indicating that courses in these clusters focus primarily on content delivery.

An interesting trend across all three clusters for HarvardX is that each cluster's courses start with a spike in reading activities. This is most likely introductory or motivational material aimed at helping students persist through the course. A similar trend can be discerned in clusters from DelftX.

While the trajectory mining approach provides insights along three structural components (length, frequency of activity types, and temporal location of activities), the transition probability approach, even though it is limited to considering a single primary structural component (transitions and the order of activities), offers concrete insights into the order of a course's activities, which makes it directly applicable to principles from the learning design literature about designing activity sequences. However, the results in Figures 7 and 8 show that the trajectory mining approach enables insightful temporal analyses in that they show how evolution of patterns and sequences over the various stages of the courses and reveal key similarities and differences not only between clusters but institutions as well.

Learning Outcomes

With service to our exploratory **RQ4**, we examine the extent to which clusters are correlated with different learning outcomes as measured by course completion rates (the proportion of verified learners earning a passing grade). To see if any of the observed differences in completion rates between clusters are statistically significant (at the $\alpha = 0.05$ level), we conducted an exploratory analysis by fitting a one-way ANOVA model. For DelftX (Figure 9 containing means and standard errors), we find the differences in neither model (transition probability and trajectory mining) to be statistically significant ($p = 0.74$ and $p = 0.31$ respectively).

For HarvardX (Figure 10), a one-way ANOVA shows that for the transition probability approach, there is a statistically significant relationship between clusters and completion rates ($p = 0.002$). We therefore conducted a Tukey post-hoc test to identify which pairs of clusters were significantly different. We observe significant differences between Clusters 1 and 5 ($p = 0.002$) and Clusters 5 and 6 ($p = 0.004$). The ANOVA model for the trajectory mining approach was not statistically significant ($p = 0.39$). We present any differences strictly as correlation (not causal) and a sign that more work should be done in the future to explore any causality in this relationship.

DISCUSSION

The selection of the two institutions for the current study was a product of both of them having offered a large number of MOOCs and a mutual interest and willingness to collaborate. While these institutions combined offer a large collection of courses, they represent less than 2% of all institutions (and less than 10% of courses) on the edX platform. More generalizeable findings are likely found by including more courses and institutions in future analyses.

Regarding the findings from **RQ4** from HarvardX, while we are not yet equipped with enough evidence to present this as a causal relationship, we note that HarvardX not only has more courses than DelftX, but also more variation in the learning design and structure of courses. We are encouraged that our methods show differences in learning outcomes based on our course-design abstraction, and this further indicates that this research would benefit greatly from the involvement of more institutions so that we can consider the full spectrum of learning designs and continue to dig deeper into their relationship with learning outcomes.

Future work should explore to what extent increasing the number of grams (sequences longer than two pairs of activities/elements) for the transition probabilities can impact the (i) insights afforded by the results and visualization and (ii) learning outcomes from each cluster. It should also be insightful if in future research, instead of taking the entire course to encode as input, one conducted a similar method using only course chapters or weeks.

Additionally, we explored the predictive power of our course design data on a course by course basis for HarvardX. In the linear model, predictor variables included total number of activities and transitions, and the outcome variables included certification rates and the percentage of chapters visited (a proxy for learner engagement). Each variable was transformed $x \to \log(1+x)$ prior to regression and normalized to unit variance. For each outcome variable we performed a step-wise regression to identify the optimal subset of predictor variables.

We found virtually no predictive power for certification outcomes using multi-regression (R^2 nearly zero), but did find significance for grade and the percentage of chapters explored ($p \leq 0.05$) within our regression coefficients for 15 of our 25 predictors. For activity frequencies, we found the number of reading activities and section breaks significant, with a negative effect on both the grade and on the percentage of chapters explored (the R^2 of the regression was 0.26 for the grade and 0.63 for the percentage of chapters explored).

We discuss this modeling simply to indicate our abstraction of course design may have predictive power for aspects of learner behavior, i.e., not just outcomes such as grades or certification. Our future work plans to address this more deeply by taking advantage of broader categories of learner metrics.

CONCLUSION

In this research we present a successful method of abstracting the design of a MOOC according to principles from the learning design literature (**RQ1**). Using this method we then quantitatively compare and contrast the design of the courses using both transition probability clustering and trajectory mining (**RQ2**). This then enabled us to draw qualitative insights about the commonalities among courses in each cluster—revealing latent themes in learning design patterns by MOOC instructors and designers (**RQ3**). To explore the validity of these findings, we evaluate the extent to which these identified trends in the learning design are associated with learning outcomes in the courses examined (**RQ4**). This new avenue of documenting and understanding pedagogy at scale enables novel lines of inquiry in online learning research by directly connecting teaching/ learning design trends to measurable trends in learner engagement.

We are inspired by our ability to automate the process of categorizing course designs and propose that future work needs to continue to refine and test our abstraction method and how it impacts categorization. We also hope to expand our outcome metrics in order to further explore the relationships with course design. Above all, we hope that our work will be a first step in showing the value of addressing digital learning environments from a course structure perspective and finding new challenges as digitization takes an even firmer hold in the learning sciences.

ACKNOWLEDGEMENTS

We thank the VPAL-Research group for access to the HarvardX pre-processed data. Ilia Ruskin and Huan Wang are specifically thanked for exploratory work in this study.

REFERENCES

1. Andrew Abbott and Angela Tsay. 2000. Sequence analysis and optimal matching methods in sociology: Review and prospect. *Sociological methods & research* 29, 1 (2000), 3–33.
2. Mina Boroujeni and Pierre Dillenbourg. 2018. Discovery and Temporal Analysis of Latent Study Patterns from MOOC Interaction Sequences. In *Proceedings of the Eigth International Learning Analytics & Knowledge Conference*. ACM.
3. Tadeusz Caliński and Jerzy Harabasz. 1974. A dendrite method for cluster analysis. *Communications in Statistics-theory and Methods* 3, 1 (1974), 1–27.
4. Dan Davis, Guanliang Chen, Claudia Hauff, and Geert-Jan Houben. Gauging MOOC Learners' Adherence to the Designed Learning Path. In *Proceedings of the 9th International Conference on Educational Data Mining*. 54–61.
5. Michael Derntl, Susanne Neumann, and Petra Oberhuemer. 2009a. Constructing and Evaluating a Description Template for Teaching Methods.. In *EC-TEL*. Springer, 447–461.
6. Michael Derntl, Susanne Neumann, Petra Oberhuemer, Colin Tattersall, Dominique Verpoorten, and Roland Klemke. 2009b. *Report on the standardized description of instructional models*. Technical Report. University of Vienna.
7. Chase Geigle and ChengXiang Zhai. 2017. Modeling Student Behavior With Two-Layer Hidden Markov Models. *Journal of Educational Data Mining* 9, 1 (2017), 1–24.
8. René F Kizilcec, Mar Pérez-Sanagustín, and Jorge J Maldonado. 2017. Self-regulated learning strategies predict learner behavior and goal attainment in Massive Open Online Courses. *Computers & education* 104 (2017), 18–33.
9. Peter Langfelder, Bin Zhang, and Steve Horvath. 2007. Defining clusters from a hierarchical cluster tree: the Dynamic Tree Cut package for R. *Bioinformatics* 24, 5 (2007), 719–720.
10. Diana Laurillard. 2008. Technology enhanced learning as a tool for pedagogical innovation. *Journal of Philosophy of Education* 42, 3-4 (2008), 521–533.
11. Diana Laurillard. 2013. *Teaching as a design science: Building pedagogical patterns for learning and technology*. Routledge.
12. Diana Laurillard and Dejan Ljubojevic. 2011. Evaluating learning designs through the formal representation of pedagogical patterns. *IGI Global* (2011).
13. Diana Laurillard and Patrick McAndrew. 2003. Reusable educational software: A basis for generic learning activities. *Reusing online resources: A sustainable approach to e-learning* (2003), 81–93.
14. Vladimir I Levenshtein. 1966. Binary codes capable of correcting deletions, insertions, and reversals. In *Soviet physics doklady*, Vol. 10. 707–710.
15. Yanchi Liu, Zhongmou Li, Hui Xiong, Xuedong Gao, and Junjie Wu. 2010. Understanding of internal clustering validation measures. In *Data Mining (ICDM), 2010 IEEE 10th International Conference on*. IEEE, 911–916.
16. Jorge Maldonado-Mahauad, Mar Pérez-Sanagustín, René F Kizilcec, Nicolás Morales, and Jorge Munoz-Gama. 2017. Mining theory-based patterns from Big data. *Computers in Human Behavior* (2017).
17. Anoush Margaryan, Manuela Bianco, and Allison Littlejohn. 2015. Instructional quality of massive open online courses (MOOCs). *Computers & Education* 80 (2015), 77–83.
18. Yishay Mor and Niall Winters. 2007. Design approaches in technology-enhanced learning. *Interactive Learning Environments* 15, 1 (2007), 61–75.
19. Patrick Mukala, JCAM Buijs, and WMP Van Der Aalst. 2015. Exploring students' learning behaviour in moocs using process mining techniques. (2015).
20. Quan Nguyen, Michal Huptych, and Bart Rienties. 2018. Linking Students' Timing of Engagement to Learning Design and Academic Performance. In *Proceedings of the Seventh International Learning Analytics & Knowledge Conference*. 141–150.
21. Zachary A Pardos, Steven Tang, Daniel Davis, and Christopher Vu Le. 2017. Enabling Real-Time Adaptivity in MOOCs with a Personalized Next-Step Recommendation Framework. In *Proceedings of the Fourth (2017) ACM Conference on Learning@ Scale*. ACM, 23–32.
22. Peter J Rousseeuw. 1987. Silhouettes: a graphical aid to the interpretation and validation of cluster analysis. *Journal of computational and applied mathematics* 20 (1987), 53–65.
23. Lee S Shulman. 2005. Signature pedagogies in the professions. *Daedalus* 134, 3 (2005), 52–59.
24. Karen Swan, Scott Day, Leonard Bogle, and Traci van Prooyen. 2014. AMP: A tool for characterizing the pedagogical approaches of MOOCs. *e-mentor* 2 (2014), 75–85.
25. Miaomiao Wen and Carolyn Penstein Rosé. 2014. Identifying latent study habits by mining learner behavior patterns in massive open online courses. In *Proceedings of the 23rd ACM International Conference on Conference on Information and Knowledge Management*. ACM, 1983–1986.