Distance Education

ISSN: 0158-7919 (Print) 1475-0198 (Online) Journal homepage: http://www.tandfonline.com/loi/cdie20

Exploring the temporal dimension of forum participation in MOOCs

Hengtao Tang, Wanli Xing & Bo Pei

DISTANCE EDUCATION
https://doi.org/10.1080/01587919.2018.1476841

Routledge
Taylor & Francis Group

ARTICLE

Exploring the temporal dimension of forum participation in MOOCs

Hengtao Tang[a], Wanli Xing[b] and Bo Pei[c]

[a]Department of Learning and Performance Systems, Pennsylvania State University, University Park, PA, USA; [b]Department of Educational Psychology & Leadership, Texas Tech University, Lubbock, TX, USA; [c]Department of Educational Psychology & Leadership, Texas Tech University, Lubbock, Texas

ABSTRACT
Discussion forums are increasingly central to massive open online courses (MOOCs), and it is vital for learners to participate in associated forum activities. Active forum participation positively relates to learner achievement in that more posts yield better learner performance. However, this numerically aggregated measure overlooks the fact that longitudinal trajectories of forum participation temporally vary among learners with different motivations in taking MOOCs. To provide timely support for learners to stay engaged, it is important to understand the temporal variation of longitudinal forum participation and how different motivations account for the variance. Using educational data mining techniques, this research identified three clusters with different longitudinal participation trajectories and also indicated that intrinsically motivated learners outperformed others in their longitudinal forum engagement. Also, examining longitudinal forum participation more accurately differentiated learner performance than the numerically aggregated measure. Last, learners persistently engaging in forums were more likely to perform better in MOOCs.

ARTICLE HISTORY
Received 9 September 2017
Accepted 11 May 2018

KEYWORDS
MOOC forum; temporal dimension; educational data mining; longitudinal trajectory; motivation; learner performance

Introduction

Discussion forums are primary social settings in massive open online courses (MOOCs) and learning taking place in these forums "holds the promise of scalable peer-based learning" (Brinton et al., 2014, p. 346). However, MOOCs have not entirely fulfilled this promise yet, largely because discussion forums have not constantly engaged learners in associated forum activities (Yang, Sinha, Adamson, & Rose, 2013). Participation in forum activities is required for learners to benefit from discussion forums (Xing, Chen, Stein, & Marcinkowski, 2016), and becomes a priority for learners to thrive in MOOCs. More attention has thus shifted to sustaining learner forum participation in these large-scale courses.

Posting and viewing are two dominant activities gauging learner forum participation and are often used to analyze learner performance (Wise, Zhao, Hausknecht, & Chui, 2013). Numerous studies have claimed a positive relationship between learner performance and forum participation in MOOCs with regard to the summative numbers of posts and views.

CONTACT Wanli Xing wanli.xing@ttu.edu Department of Educational Psychology and Leadership Institution, Texas Tech University, 3008 18th Street, Lubbock, TX, USA

For example, forum participants with more posts and views are more likely to outperform their inactive counterparts in MOOCs (Anderson, Huttenlocher, Kleinberg, & Leskovec, 2014; Engle, Mankoff, & Carbrey, 2015). Nevertheless, a concern is raised about whether the converged manner used to quantify forum participation patterns influence the effectiveness of findings about how participation influences learner performance (Bergner, Kerr, & Pritchard, 2015). Some significant relationships between forum participation and learner performance in the earlier weeks discontinue and even reverse in later weeks. For example, using week-by-week analysis rather than summative measures, Zhu et al. (2016) found that the way learners connected with peers in the forum activities during Week 2 was correlated to their scores in weekly quizzes, but this effect became invalid in subsequent weeks. In short, the validity concern about summative measures primarily results from the fact that these summative measures overlook a key dimension of learner forum participation – time (Goggins & Xing, 2016).

Time is a significant topic of learning (Barbera, Gros, & Kirschner, 2015; Reimann, 2009). For Reimann (2009), learning unfolds over time as a cumulative process of interdependent events, so the time sequence of these events taking place significantly influences the learning process. Without considering the temporal dimension, researchers cannot uncover subtle details that influence learning (Kizilcec, Piech, & Schneider, 2013; Molenaar, 2014), such as the temporal variation of forum participation across a longitudinal trajectory (Kapur, Voiklis, & Kinzer, 2008). In online settings, learner participation patterns fluctuate during different phases and for various barriers (Liu, Wang, & Tai, 2016; Wise et al., 2013), so understanding the temporal variation is required to provide timely scaffolding to constantly engage learners in forum activities. For example, some actively engaged learners in the first several weeks of MOOCs gradually disengage in later weeks (Yang et al., 2013). If timely scaffold were available, these learners might have sustained their engagement in forum activities. Taking time into consideration is critical to obtain a substantial view of online learner forum participation and to help them maintain forum participation in MOOCs.

In addition to understanding the temporal variation of forum participation, it is also necessary to identify the factors that influence temporal learner forum participation in MOOCs. Participation is a longitudinal trajectory of engagement affected by various motivational factors (Pintrich & Schrauben, 1992). For example, participation trajectories differ between learners with intrinsic and extrinsic motivations of participating in the course. Intrinsic motivation involves "an inherent gratification prompted by the feeling that learning is interesting and enjoyable" (Barak, Watted, & Haick, 2016, p. 50), but extrinsic motivation drives learners with external rewards and incentives (Pintrich & Schrauben, 1992). Accordingly, intrinsic motivation is closely associated with active participation as intrinsically motivated learners are generally interested in the course and confident in performing well (Pintrich & Schrauben, 1992). However, others with extrinsic motivation tend to find it difficult to stay longitudinally engaged (Csikszentmihalyi, 1975; Pintrich & Schrauben, 1992). MOOCs enable thousands of learners with different intentions to enroll in desired courses. Students with extrinsic motivation might not be as determined as their intrinsically motivated peers to overcome various barriers and thus their participation pattern fluctuates. This confirms the importance of linking learners' different longitudinal participation trajectories with their individual motivations. In this way, we can capture learner motivation at an early stage of MOOCs to provide efficient timely scaffolds for disengaged learners.

The purpose of this study was thus threefold. First, using educational data mining techniques, this research explored the temporal evolution of learner forum participation and revealed different longitudinal trajectories of forum participation. Second, the research sought to identify how different learner motivation influenced their longitudinal forum participation. Third, the research investigated the relationship between longitudinal trajectories of forum participation and learner performance in MOOCs. The research findings are significant because this research adds the time dimension into the effort to reinvigorate learner forum participation and reveals more subtle details of learner forum behaviors in MOOCs (Kizilcec et al., 2013). These details can not only help researchers understand the complexity of learner forum participation in this setting but also enable us to propose relevant strategies for educators to sustain learner forum participation and to improve their performance.

Literature review

Time and the temporal dimension of learning

Time is a fundamental dimension of human development and educational practices (Vygotsky, 1978) but remains underexplored (Barbera et al., 2015; Reimann, 2009). For Reimann (2009), this is because the unit of analysis in educational research is misunderstood. Traditional research regard variables as the unit of analysis; thus, its primary purpose is to address the causal relationship between the independent and dependent variables by observing the variance of value from the input to the output of learning (Molenaar, 2014). In this variable-centered view, the causal relationship is of temporal homogeneity during the course of learning, assuming that learning is a static artifact (Kapur et al., 2008). However, when breaking down a course into weeks, researchers found that existing relationships between variables might not continue across the learning process, either disappearing or reversing at some points (Zhu et al., 2016). For a solid understanding of learning, making sense of these temporal variations is critical. Otherwise, the explanatory power of the variable-centered research is subject to uncertainty (Molenaar, 2014). Therefore, it is necessary to understand the learning process from an alternative perspective – an event-based view.

The event-based view of the learning process takes temporality into consideration, assuming that learning is an inherently cumulative process of events (Mercer, 2008; Reimann, 2009). The unit of analysis in this view is an event. This event-based view has advantages over the former perspective in three aspects. First, this view considers the learning process as a “developmental event sequence, not a change in values of process variables” (Reimann, 2009, p. 247). In this way, each event of the longitudinal trajectory matters for learning, rather than in a variable-centered view which only considers the input and the output states of variables. Second, this event-based view provides a unique perspective to understand the temporal variation of learner participation across the learning process. During the course of learning, the trajectory of transitions between events results in fluctuating patterns of learner participation. By attending to a longitudinal trajectory of events, researchers can address the temporal variation of participation over time. Third, the event-based view looks into the impact of the time sequence on learning (Reimann, 2009). Research regarding temporality focuses mainly on time efficacy, time use, and time pace, but overlooks time sequence (Barbera et al., 2015). In the event-based view, the time sequence of each event determines that events taking place at different time points have distinct impacts on learning outcomes. For example, learners who initially post infrequently but submit plentiful forum posts right before the final exam usually cannot perform well in the course (Canal, Ghislandi, & Micciolo, 2015). In sum, the event-based view outperforms the former one in revealing more subtle but significant details about learning by considering the time issue. To fully utilize the benefits of the event-based view, the longitudinal trajectory of the learning process is the recommended focus for future research (Barbera et al., 2015; Canal et al., 2015).

Research regarding online forums has also confirmed the advantage of focusing on the longitudinal trajectory of forum participation in offering pertinent scaffolds. Canal et al. (2015) indicated that high performing students tend to maintain more active forum participation than others throughout the course. They recommended that online instructors should help learners maintain regular forum participation rather than a larger number of total forum access (Canal et al., 2015). This implication might be an alternative of summative measures to reinforce learner forum participation in MOOCs. In addition, Kizilcec et al. (2013) more accurately portrayed how different clusters of learners interact with MOOCs by addressing their longitudinal engagement trajectory. It is worth noting that Kizilcec et al. (2013) excluded learner forum participation patterns from the clustering algorithm, but their subsequent correlational analysis specified participation patterns in forum activities as a unique measure to differentiate longitudinally active participating students and any other inactive clusters. Consequently, it is important to investigate this temporal perspective and seek more substantial implications of forum participation in MOOCs.

Factors influencing learner forum participation in MOOCs

Empowering efficient discussion forums has become increasingly vital to the success of MOOCs. Discussion forums are the main place for social interactions in MOOCs. For learners, interactions taking place in forums allow them to offset the lack of instructional support (Zhang, Skryabin, & Song, 2016). Active forum participation is important for learners, but the overall volume of forum participation drops steadily throughout the course offering (Brinton et al., 2014). A majority of learners reduce and even terminate forum participation before the course ends. In particular, many learners intending to complete the course gradually disengage from discussion forums and quit the course. This hurdle of constantly engaging learners in the forum activities has become a major problem of MOOCs (Yang et al., 2013).

Understanding the factors influencing learner forum participation is necessary to tackle this problem. Reinforcing learner forum participation has been a central topic of online learning research (Hampel & Pleines, 2013; Thomas, 2002; Xie, DeBacker, & Ferguson, 2006). Traditional online courses attempt to include forum participation in the course assessment (Hampel & Pleines, 2013), but conversely, many learners casually post numerous superficial items to meet the requirement for passing the course (Thomas, 2002). This type of forum participation is the behavioral outcome of extrinsic motivation towards completing required coursework (Thomas, 2002). In contrast, intrinsically motivated learners who perceive the course as relevant and enjoyable are more actively engaged in forum activities (Xie et al., 2006). However, as the course continues, their level of intrinsic motivation slumps progressively and thus their forum participation also declines (Xie et al., 2006). In short, the type of learner motivation significantly influences the way online learners participate in forum activities.

Research has also indicated that learner motivation is a remarkable predictor of forum participation patterns in MOOCs (Kizilcec & Schneider, 2015; Xiong et al., 2015). That is, motivated learners are more likely to overcome various barriers and maintain persistent participation than those without motivation (Xiong et al., 2015). Although MOOCs differ from traditional online courses in that learners can self-determine forum participation disregarding penalties and costs, choice of time and effort invested in forum activities is an expression of learner motivation (Kizilcec & Schneider, 2015). For example, Huang, Dasgupta, Ghosh, Manning, and Sanders (2014) investigated the forum participation patterns of "superposters" in one course and also across Coursera datasets. They speculated that superposting was an intrinsic trait of these active learners because they contributed to the forum interactions with high-volume, high-quality, and high-impact posts in one course. In addition, this group of learners tended to register for more courses and also were more likely to be superposters in other enrolled MOOCs (Huang et al., 2014). It seems that intrinsic motivation was more relevant to active forum participation. In addition, Kizilcec and Schneider (2015) identified several types of extrinsic motivation that were also related to the increase of forum participation patterns. Specifically, learners who hoped to meet new people or intended to earn certificates were more actively engaged in forum discussions. In contrast, learners who took courses with their friends or as a part of their research or courses had a higher probability of infrequent forum participation.

In all, the existing evidence fails to provide a unified understanding of whether intrinsic or extrinsic motivation influences learner forum participation. This dilemma might result from overlooking the temporal dimension of forum participation because the sole reliance on summative measures might undermine the explanatory power of these investigations (Molenaar, 2014). For example, Huang et al. (2014) recognized superposters in MOOCs using their summative volume rather than the longitudinal trajectory of forum participation. Adding to this, whether and how the type of motivation affects learner forum participation trajectory remain unknown. It thus requires further investigations to obtain a more consolidated understanding of this complexity regarding how learner motivation shapes longitudinal forum participation in MOOCs.

Forum participation and learner performance in MOOCs

Research suggests that learner participation has the most significant influence on learner performance (Phan, McNeil, & Robin, 2016). Learners actively engaging in MOOCs are more likely to outperform their peers without this trait. In particular, encouraging learner participation in forum activities is integral to help learners complete a MOOC (Anderson et al., 2014).

The significance of forum participation for learner performance in MOOCs has been widely evidenced. For example, Anderson et al. (2014) and Engle et al. (2015) claimed active forum participants normally outpaced others in assignments and quizzes and those students were more likely to complete a MOOC. However, these arguments build upon the summative measures of forum participation. Given the impact of temporality, the relationship between learner behavior in forums and their performance is not constant but varies temporally (Zhu et al., 2016). In addition, participation at different time points makes distinct contributions to learner performance (Barbera et al., 2015; Canal et al., 2015; Rosé et al., 2014). For example, Rosé et al. (2014) reported that learners who had been actively involved in forum discussions since the beginning were less likely to quit. On the other hand, abruptly increased forum posts right before the final exam seldom helped initially inactive learners earn a high grade in online courses (Canal et al., 2015). Taking time into consideration provides a better view of how learning actually unfolds but also potentially reverses the existing conclusions drawn from summative measures (Molenaar, 2014). Given that most research on forum participation is geared toward the improvement of learner performance in MOOCs, it is thus necessary to re-examine the relationship between temporal forum participation and learner performance in the event-based view.

Methodology

Datasets

The dataset used in this research was derived from a MOOC offered by a college in the northeastern United States. The MOOC was launched in August 2014 on Canvas Network. The primary emphasis of this MOOC was to introduce the concepts, techniques, and principles of project management. The course lasted for eight weeks and included 11 modules. Each module came with an exclusive forum and an online quiz. Participation in these forums was voluntary, and the learners revisited prior forums at will. Further, 12 quizzes including a final quiz were given across the course. Upon completion of the course, participants earned a certificate that could be used to meet professional development requirements for the Project Management Professional certification issued by the Project Management Institute.

The dataset used in this research was mainly retrieved from two sources: clickstream data and the Canvas application programming interface (API). First, we collected the clickstream data for the duration throughout the course from Canvas. The clickstream data represented a specific collection of pages that learners had viewed with information available on when or how many times they viewed those pages (e.g., which quiz they completed, how many times they viewed this quiz, how many views they had on a discussion forum and when for each view). The other source of data was obtained through the Canvas API; this represented mainly the information about quiz scores and the discussion forum data. This part of the dataset was in the JSON format (https://www.json.org/) and comprised each learner's quiz scores and their longitudinal activities in the discussion forum (e.g., who posted in the discussion forum and when for each post). In the end, 607 learner forum activity records were kept for this analysis.

Variables

Forum participation

This study considered learner participation in MOOC discussion forums using a week-by-week analysis (Bergner et al., 2015). Forum participation was characterized by attributes including the number of forum posts, forum comments, and forum views per week.

Performance

Learner performance was described as the average grade that each learner earned in all quizzes in this MOOC.

Data processing and analysis

Before data analysis, the dataset was normalized to mitigate the difference in the scale between variables associated the with forum participation of one participant (Xing, Guo, Petakovic, & Goggins, 2015). The result of the longitudinal trajectory after the normalization between 0 and 1 is visually represented in Figure 1.

Table 1 reveals the basic descriptive statistics about learner forum participation. As the table shows, the weekly forum participation of each learner in the discussion forum decreased from 6.11 to 2.87, with a minor increase from Week 4 to Week 5 and from Week 7 to Week 8. Furthermore, there was a decreasing variance among learner forum participation as time proceeded.

Table 2 presents an overview of the methodology implemented in this study. A description of more detailed procedures of data analysis will follow.

The primary data analysis method used in this study was the longitudinal k-means clustering algorithm (KmL). KmL is an unsupervised educational data mining technique grouping participants by the longitudinal trajectory of their temporal profile rather than the numerical convergence of the mean levels (Genolini & Falissard, 2010). Despite its infrequent use in educational research, KmL has been widely used in research in medicine and psychiatry (Pingault et al., 2014, 2011). As a simpler and more robust method to cluster participants using the temporal data, KmL fits well with the purpose of this study of addressing the temporal dimension of forum participation.

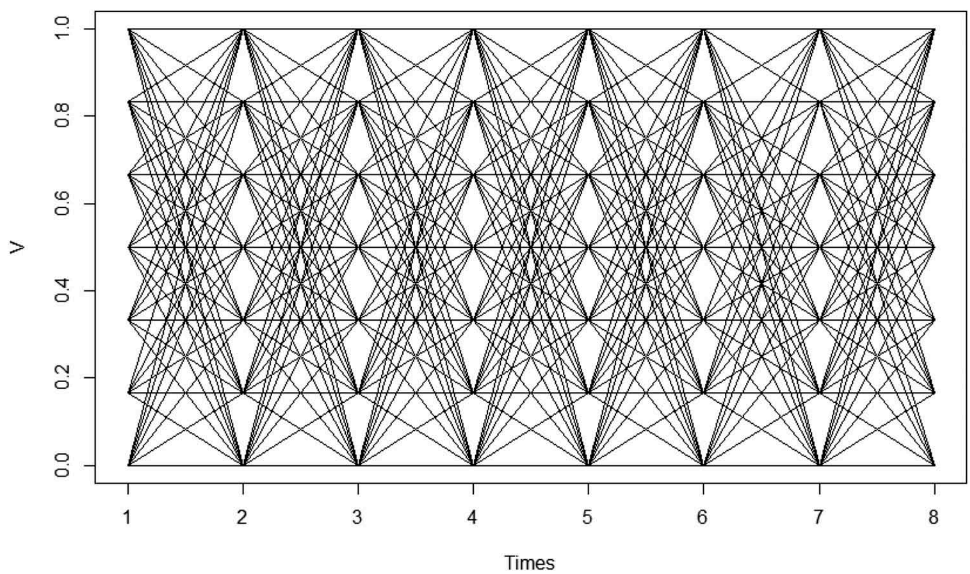

Figure 1. The forum participation trajectory was between 0 and 1 after the normalization.

Table 1. Descriptive statistics of learner participation in the discussion forum.

Week(s)	Mean	Range	SD
1	6.11	118	11.24
2	4.44	126	10.15
3	3.15	100	7.45
4	2.61	68	5.90
5	2.78	70	6.62
6	2.32	47	4.98
7	2.10	47	4.55
8	2.87	51	6.04

Table 2. Data analysis methods used in this research.

Research purpose(s)	Data analytic method(s)
Identify the longitudinal trajectory of learner forum participation in MOOCs	Longitudinal k-means clustering algorithm; pseudo-T square test; ISODATA algorithm
Identify the relationship between learner motivation and their forum participation trajectory	Constant comparative method; Chi-square analysis
Identify the relationship between learner forum participation trajectory and their performance	Analysis of variance (ANOVA); Tukey Honestly Significant Difference (HSD) tests

The first step was to determine an optimal number of clusters that would be used in this research. To increase the reliability, this research ran the pseudo-T square analysis (Duda & Hart, 1973) and the iterative self-organizing data analysis technique (ISODATA) algorithm (Ball & Hall, 1965) as a two-step verification (assuring two values matched).

The pseudo-T square analysis identifies the optimal clustering number based on the within-cluster dispersion, W_k, generated when data was grouped into K clusters (Tibshirani & Walther, 2005). As the number of clusters increases, the value of W_k would decrease during which W_k has the most dramatic fluctuation at a point. This point indicates the optimal number of clusters (Tibshirani & Walther, 2005).

The ISODATA algorithm is a classic statistical method that runs the algorithm iteratively until the threshold value is obtained. Using K as the number of clusters, the ISODATA algorithm computes the total value of the within-cluster sum of squares (WSS) for each K-value (e.g., 2–10 in this research) and plots the curve of how WSS changes. In general, the optimal value of K is seen where the curve shows an apparent bend.

Assuring two values of K from the two-step verification matched, the researchers then applied KmL to process the data (Genolini & Falissard, 2010). The researchers assigned a total of 607 participants into the optimal *K* clusters using their longitudinal trajectory of forum participation patterns (*X*) across eight weeks (see Figure 2). The KmL algorithm used in this study refers to the Euclidean distance [Dist (X_n, X_m), see equation (1)], which shows the degree of similarities between the two participants (Genolini & Falissard, 2010). Taking the example of two participants (*n* and *m*), a smaller value of the Euclidean distance indicates they have more similarities and are also more likely to be in the same cluster.

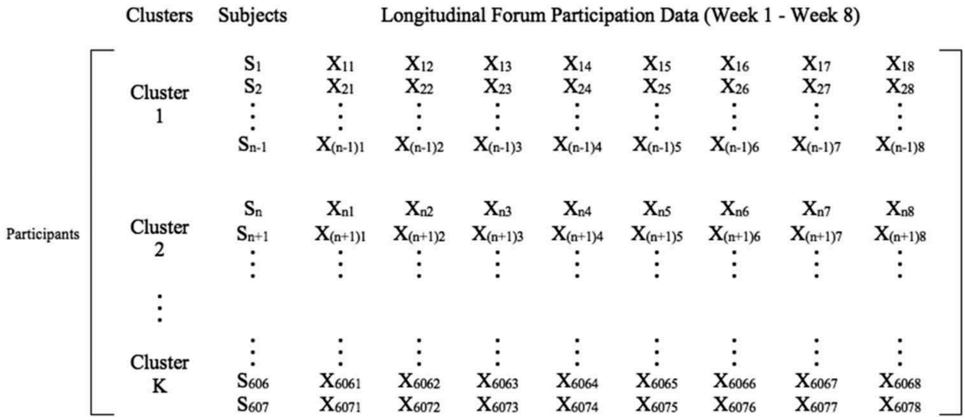

Figure 2. Flow chart for clustered longitudinal forum participation data.

$$\mathrm{Dist}(X_n, X_m) = \sqrt{\frac{1}{T}\sum_{t=1}^{T}(X_{nt} - X_{mt})^2} \tag{1}$$

The next step was to identify how learner motivation influences their longitudinal trajectory of forum participation. To obtain the motivation data, two researchers processed all the self-introductory forum posts. Coding these self-introductory posts revealed information regarding their statement of motivations. It was worth noting that not all participants posted in this forum. Three hundred learners composed a total of 444 posts in the introductory forum, but 144 comments not relevant to learner motivation were removed before coding.

The constant comparative method was applied to code these posts (Glaser & Strauss, 1967). For the coding process, each introductory post by each participant was the unit of analysis. Learner motivation for taking this course was the coding scheme (see Table 3). In particular, learner motivation was differentiated into extrinsic and intrinsic motivation based on the strategic motivation framework (Pintrich & Schrauben, 1992). To ensure research rigor, each researcher coded 30 posts (10% of the overall posts) individually to obtain the inter-rater reliability of the coding completed by the first author. Cohen's (1960) kappa was calculated and reached an acceptable level ($k = 0.83$) in this research. Then the first author continued to code all of the remaining forum posts (Landis & Koch, 1977). However, some participants did

Table 3. Coding schemes of self-introductory forum posts in the first week.

Codes	Excerpts from forum posts
Extrinsic motivation	I am a registered nurse and midwife with over 25 years experience and I am also an artist. I am hoping to move into Project Management at work, so this course is a way to learn and update my skills.
	I'm currently working as a Project Coordinator but I would like to get more knowledge on the basic principles of Project Management to further my career.
Intrinsic motivation	I have no formal exposure to the PM. It sounds interesting and this course will increase my knowledge base about subject matter.
	Currently working as a PM in low voltage commercial construction. Lots of experience with the industry and our installations but need a better grasp on the professional side of the business.
N.A.	Nice to meet u all!!! Feel free to text me anytime if something is needed.

not clarify their motivation and thus these posts were removed. In the end, 268 posts were kept for further analysis. Once the coding procedure was done, the researchers referred to the user identification to match each post to the participant who wrote it. The researchers administrated a chi-square test to identify the significance of the difference in motivation among the three clusters.

The last analysis was conducted to identify the relationship between learner longitudinal forum participation and their course performance. For this step, the researchers applied the ANOVA and the Tukey HSD posttest to detect any significant performance differences among the three clusters.

Results

What is the longitudinal trajectory of learner forum participation in MOOCs?

To determine the optimal number of clusters, the pseudo-T square analysis and the ISODATA algorithm were conducted as two-way verifications. The result revealed the optimal number of clusters in this research was three (see Figures 3 and 4). Therefore, the research categorized participants with different longitudinal traits of forum participation into three clusters. Figure 5 visually exhibits the forum participation trajectories of these three clusters as solid colored lines. Table 4 provides descriptive statistics results associated with each cluster.

Cluster A (seldom engaging) – 47.3% of learners were in this cluster, which showed an overall trajectory of low forum engagement. Despite a slight increase from Week 2 to Week 5, their overall forum participation was not satisfying. Additionally, the average participation

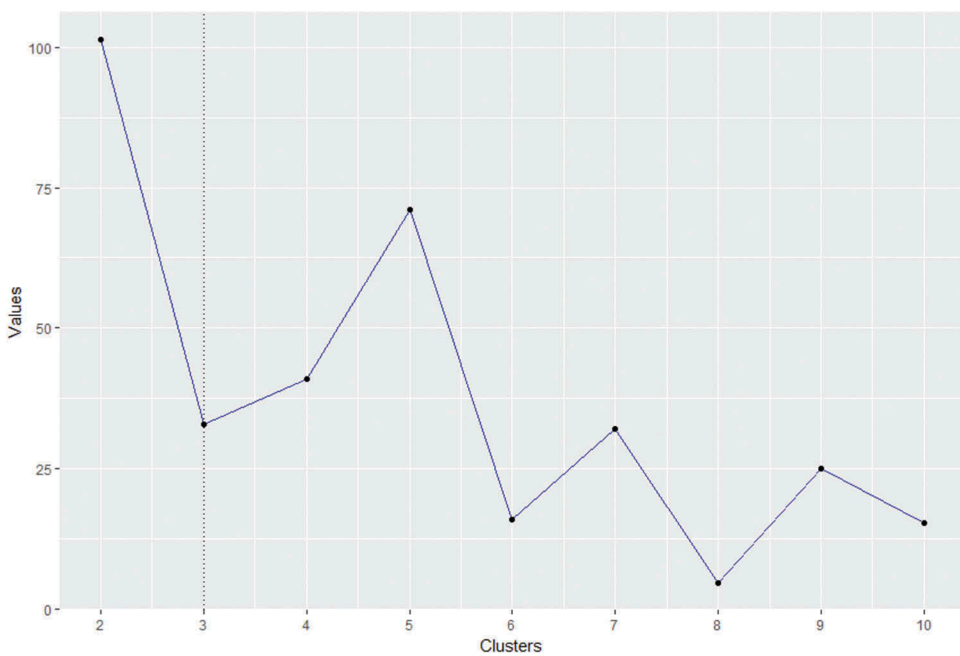

Figure 3. Pseudo-T square results showing the optimal clustering number is three.

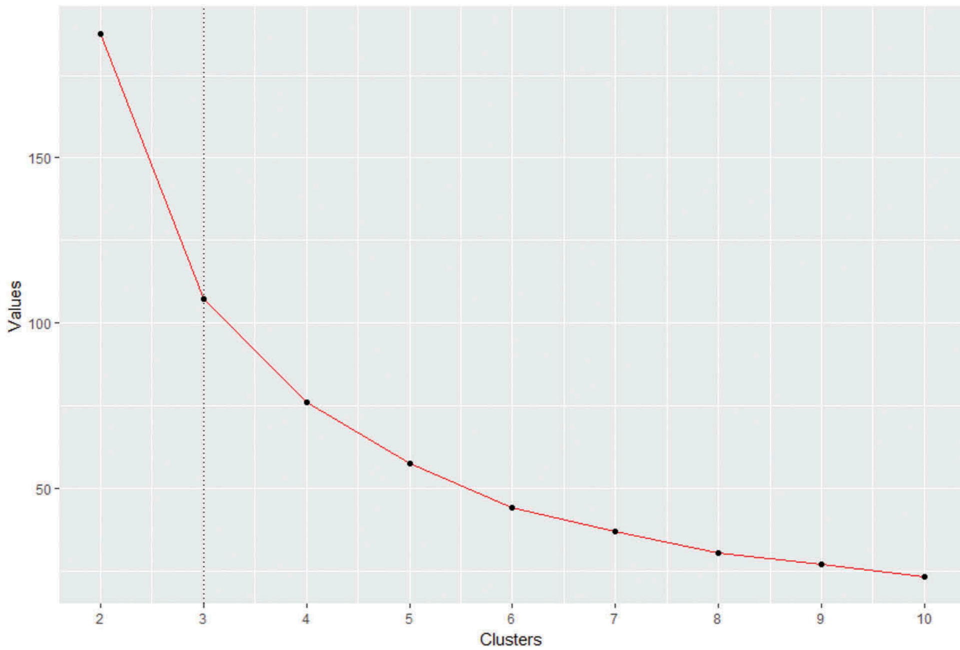

Figure 4. ISODATA algorithm results showing the optimal clustering number is three.

of this cluster rose to the highest level during the last week. Some participants in this cluster actively contributed to the forum of this week (max = 45) probably due to the incentive of the final quiz and the expectation of receiving certificates.

Cluster B (gradually disengaging) – 36.2% of participants were in this cluster, which demonstrated a significant decrease in forum participation across weeks. At the beginning, this cluster of learners was highly engaged in discussion forums and several of them even topped the list as the most active participants (max = 118), but overall, their patterns in the first week had a large variance. The following weeks saw a significant decrease in their averaged forum participation as many initially active participants became gradually disengaged. The decreased variance of forum participation meant most participants in this cluster had maintained low forum participation since then. It is worth noting that some initially active participants subsequently withdrew from the forum discussion. These learners had a large volume of forum participation patterns in the first week. From a summative perspective, they might be superposters, but more accurately, they should be in this gradually disengaging cluster with regard to their overall participation trajectory. Furthermore, similar to Cluster A, forum participation of this cluster rose during the last week but never returned to its highest level.

Cluster C (persistently engaging) – 16.5% of learners were in this cluster, which sustained a much higher level of forum engagement throughout the course. For Cluster C, forum participation each week was higher than any other group during the same period, except for the first week when several learners in Cluster B led in participation. Their lowest forum participation was seen in Week 7, but this number was still higher than that of the other two groups. In addition, a minor increase in forum participation occurred during the last week, resembling the other two clusters.

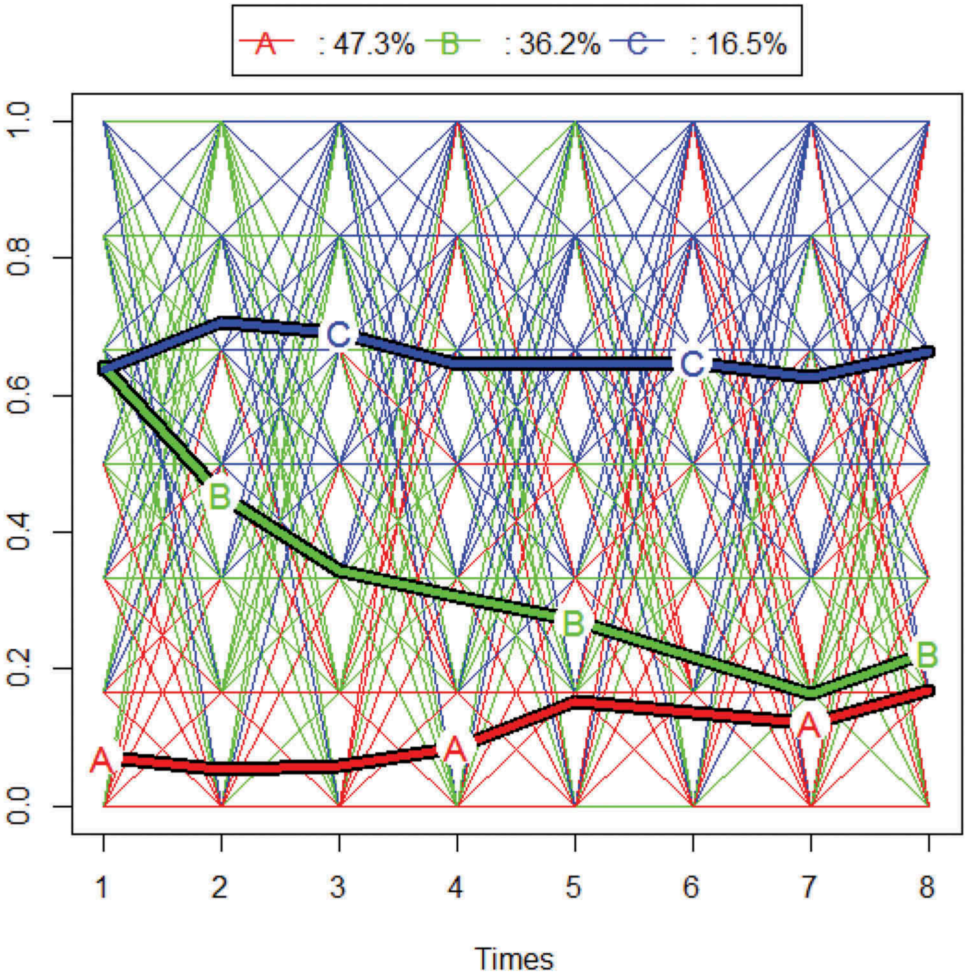

Figure 5. Three clusters of learners based on their longitudinal forum participation.

Table 4. Learner forum participation in different clusters.

	A			B			C		
Week	Mean	Range	SD	Mean	Range	SD	Mean	Range	SD
1	0.58	19	1.92	8.43	118	10.34	16.33	90	17.86
2	0.32	10	1.13	4.64	35	5.43	15.65	126	19.86
3	0.44	18	1.72	2.67	26	3.50	11.96	100	14.48
4	0.59	16	1.66	2.40	30	3.52	8.87	68	11.35
5	1.11	20	2.26	1.97	15	2.69	9.45	70	13.70
6	0.96	14	1.85	1.35	14	1.76	8.49	47	9.59
7	0.93	18	2.18	1.12	17	1.71	7.72	47	8.40
8	1.55	45	4.05	1.58	17	2.20	9.64	51	10.68

What is the relationship between learner motivation and forum participation trajectories?

To answer this question, the researchers analyzed the self-introductory forum posts in the first week. Table 5 shows the frequency of each condition shown in the dataset. Then a chi-square

Table 5. Descriptive statistics for the coded forum posts in the first week.

Clusters	Motivation		Total
	Extrinsic	Intrinsic	
A	49	15	64
B	73	51	124
C	29	51	80
Total	151	117	268

test was conducted. The result (see Table 6) confirmed that the difference in the type of motivations was significant among three clusters ($\chi^2 = 24.090$; $df = 2$; $p < .001$).

What is the relationship between the longitudinal trajectory of forum participation and learner performance?

The research conducted an ANOVA test and a Tukey HSD posttest to investigate the relationship between learner forum participation trajectory and their performance. Some participants skipped the quizzes and their performance records were removed for the analysis (see Table 7 for the descriptive statistics about learner performance). The ANOVA test result (see Table 8) showed that there was a significant difference in course performance among these clusters ($F = 79.26$, $p < .000$). That meant the performance of learners with a different longitudinal trajectory of forum participation varied significantly.

To further identify where the differences existed, the Tukey HSD posttest was conducted (see Table 9). The result revealed that the persistently engaging cluster (Cluster C) had the best course performance, followed by the gradually disengaging cluster (Cluster B) and then the seldom engaging cluster (Cluster A). Learners with consistently higher forum engagement were more likely to perform well in the MOOC. For learners with an initially active forum engagement but who gradually disengaged from the forum, their performance was in the middle of the three clusters. In other words, a large volume of forum participation patterns is inadequate to guarantee acceptable learner performance in MOOCs. A solid example seen in this study was that one participant in Cluster B had intensive forum engagement in the first week (max = 118) but his/her overall performance was at a lower level with the steady decline of forum participation in the following weeks. Likewise, for participants in Cluster A, abruptly

Table 6. The chi-square test result.

	Value	*df*	Asymptotic significance (2-sided)
Pearson Chi-square	24.090[a]	2	<.001
Likelihood ratio	24.753	2	<.001
Linear-by-linear association	23.837	1	<.001
N of valid cases	268		

[a] 0 cells (0.0%) have expected count less than 5. The minimum expected count is 27.94.

Table 7. Average grades of the course quizzes in the three clusters.

Cluster	Mean	Range	SD
A (n = 267)	2.57	15.42	4.44
B (n = 230)	6.69	15.42	5.65
C (n = 91)	9.55	15.42	5.24

Table 8. ANOVA results show a significant difference in average grades among groups.

	SS	df	MS	F	Sig.
Clusters	4074	2	2037.2	79.26	$p < 0.001$***

Table 9. Tukey HSD posttest results.

		95% family-wise confidence level		
Clusters	Mean difference	Lower bound	Upper bound	Significance
B-A	4.117973	3.046297	5.189649	$p < 0.001$***
C-A	6.978151	5.532145	8.424156	$p < 0.001$***
C-B	2.860178	1.384904	4.335452	$p < 0.001$***

increased forum participation in the last week barely helped them gain a better course performance. In conclusion, the longitudinal trajectory of learner forum participation significantly influenced their performance in MOOCs.

Discussion

This research considered the temporal dimension of learner forum participation in MOOCs by examining the temporal variations in its longitudinal trajectory. Using educational data mining techniques, the research identified three clusters of learners with different longitudinal trajectories of forum participation. To provide implications for efficient timely scaffolds, this research further investigated how the longitudinal trajectory of forum participation was determined by student motivations and also how it influenced learner performance in these large-scale courses. In particular, learners with different types of motivations maintained a distinct longitudinal trajectory of forum participation; intrinsically motivated learners were more likely to longitudinally engage in forum activities. Furthermore, looking into the longitudinal trajectory enabled a finer picture of the relationship between forum participation and learner performance (Canal et al., 2015); thus, more constant trajectories of active forum participation yielded better performance in MOOCs. In sum, taking temporality into consideration provides a substantial view of the complexity of learner forum participation (Bergner et al., 2015; Kizilcec et al., 2013).

In addition, this research fills the gap of a unified understanding of whether intrinsic or extrinsic motivation influences learner forum participation in MOOCs (Kizilcec & Schneider, 2015; Xiong et al., 2015). This research confirms that intrinsically motivated learners have a higher tendency to maintain a constant trajectory of active forum participation. Intrinsically motivated learners are more interested in and have more confidence in excelling in this course (Csikszentmihalyi, 1975; Pintrich & Schrauben, 1992; Xie et al., 2006). With the feeling that learning is interesting and enjoyable, they are more likely to overcome numerous barriers and actively contribute to the forum discussions. In contrast, extrinsic motivation makes limited contributions to improving online learner performance. For example, the abrupt increase of forum participation patterns during the last week did not ensure learners had a favorable performance. This refutes the idea of including more weekly assessment-oriented activities to boost learner forum participation (Hampel & Pleines, 2013; Thomas, 2002). On the other hand, considering temporality also provides us with an alternative view of how student motivation persists. We speculate that intrinsic motivation is not a constant variable but a similar construct to participation, which temporally varies throughout the course. For example, a considerable portion of learners in Cluster B also obtained intrinsic motivation towards the course, but their forum participation decreased gradually. Our guess is that these learners' intrinsic motivation might discontinue as time proceeded, and then this motivational change eventually led to fluctuating patterns of learner forum participation. Such speculation is built upon the findings that intrinsic motivation decreases in natural settings, but if efficient scaffolds are provided, learners' intrinsic motivation and engagement would persist (Liu, Chen, Lin, & Huang, 2017; Liu, Lin, Deng, Wu, & Tsai, 2014; Liu et al., 2016). For example, Liu et al. (2017) adopted a remix-oriented approach that was effective in maintaining learners' intrinsic motivation. In their experiment, students were allowed to remix original thoughts with existing online resources to create new artifacts. This creative endeavor provided new implications for the design of future MOOCs.

This study also furthers our understanding of the relationship between forum participation and learner performance. Research has attributed learner forum participation in MOOCs as an indicator of their performance in that more forum posts and views lead to better learner performance (Anderson et al., 2014; Engle et al., 2015). This study confirms the significance of forum participation for MOOC learners, and more importantly, extends the idea that what matters for learner performance is a constant trajectory of regular participation in forum activities rather than a large number of forum posts only in the first week or only several weeks before the final exam. This finding also echoes a longitudinal investigation on online forums that suggests online learners with the best performance are always more actively engaged in forum activities (Canal et al., 2015). Our interpretation of this finding originates in the event-based view of the learning process (Reimann, 2009). In this view, participation at a certain time point is an event, and each event contributes to learning. Furthermore, learning is a cumulative process of events and the sequence of events matters (Mercer, 2008; Reimann, 2009). That means, the "current" event is influenced by another one taking place earlier and also determines the later one. In this way, the negative impact of decreased participation at one point is transferrable to the following trajectory, even leading to attrition in a gradual manner. For educators, it is thus important to realize that learner dropout might result from the cumulative effect of an antecedent event. They are also advised to identify that moment and then offer timely support for disengaged learners.

Furthermore, this research provides empirical recommendations for online educators to efficiently improve learner performance in MOOCs and other online settings. First, a practical implication for educators is to maintain a constantly active forum participation trajectory as the objective of course facilitation. Longitudinal trajectories of forum participation are more accurate to assess learner performance. For example, some initially actively engaged forum participants in Cluster B were superposters in terms of summative measures, but actually they were in the gradually disengaging group if the temporal dimension was considered. To help learners maintain active forum participation trajectories, offering timely scaffolds at some key moments is necessary. Educators might consider hosting a synchronous video conference at the end of the first week to maximally resolve learners' concerns (Yousef, Chatti, Schroeder, & Wosnitza, 2014). Second, educators should recognize learner motivation at an early stage of online courses and then help learners develop and maintain intrinsic motivation throughout the course. A pragmatic approach is to include in the online course design a pre-survey that asks about learner motivation in addition to their demographics. On top of that, educators might integrate inquiry-based projects (Liu et al., 2017, 2014) rather than assessment-oriented activities (Hampel & Pleines, 2013) to improve learner perception of intrinsic motivation and engagement. Third, this research reiterates the significance of supporting the low-hanging fruit for the success of MOOCs (Xing et al., 2016; Yang et al., 2013). Learner performance of Cluster B becomes a key determinant of whether a massive number of learners can thrive in MOOCs. Cluster C, the persistently engaging group, is the ideal state for a MOOC learner, but some learners in Cluster B might also have succeeded in MOOCs if more temporal support had been made available at pivotal points (Yang et al., 2013; Zhu et al., 2016).

From the methodological perspective, this research also contributes to existing literature. First, this research confirms the viability of the longitudinal K-means clustering algorithm (Genolini & Falissard, 2010) as a promising method in educational data mining, especially for assessing learner longitudinal engagement or achievement. Second, this research reinstates the significance of adding temporal considerations to the algorithm in the attempt to uncover learner forum behaviors. Kizilcec et al. (2013) exclude forum activities from the clustering algorithm used in their inquiries, but this research confirms the significance of temporal considerations in the educational data mining investigations on forum participation.

The research makes substantial contributions to existing literature about learner temporal participation and performance in MOOCs. However, several limitations need further elaboration to prepare for future research. First, the research dataset was collected from one MOOC offered by one platform, which limits the generality of the findings. Second, the researchers manually coded the self-introductory posts to identify learner motivation. However, not all participants posted in the first week. The manual coding also affected the reliability of interpretations. Third, the research focused on longitudinal patterns of learner forum participation during the course offering, but a study on post-course forum participation is needed for a complete trajectory of forum participation, since Campbell, Gibbs, Najafi, and Severinski (2014) identified learners were still actively using the discussion forum in an archived MOOC. Fourth, the research mainly focused on the quantity of forum participation, but scholars argue the lack of qualitative analysis in learner forum posts might not indicate the quality of learner interaction in discussion forums (Zhang et al., 2016).

Conclusions

This explorative study contributes to the effort to reenergize a constantly engaging discussion forum in MOOCs. First, using the KmL algorithm, this study identifies distinct longitudinal trajectories of learner forum participation in MOOCs. This finding offers new perspectives to provide timely support for learners to stay engaged in MOOCs. Second, the research reveals that intrinsically motivated learners are more likely to longitudinally engage in MOOCs. Instead of assessment-oriented activities (Hampel & Pleines, 2013), this study recommends that educators foster the intrinsic motivation of learners by allowing them to work on inquiry-based, hands-on projects. Third, this study confirms that learners who persistently engage in discussion forums over time are more likely to excel in MOOCs. Taking time into consideration, this study reiterates that longitudinal trajectories of forum participation are more accurate than summative convergence to assess learner performance in MOOCs. Finally, the research validates the longitudinal K-means clustering algorithm (Genolini & Falissard, 2010) as a promising method in educational data mining.

To further support MOOC learners in a timely manner, future research might extend the existing investigation using a larger dataset of different MOOCs from various platforms to identify pivotal moments that determine learner performance (Wise et al., 2013). Additionally, future research might adopt a more elaborated time frame with considerations of post-course forum activities in archived MOOCs. Moreover, future research might address content-related discussion threads in MOOCs and investigate learner forum participation from both quantitative and qualitative perspectives.

Acknowledgments

The research described in this article originated from a proposal (that was later funded) submitted to Penn State EdTech Network Initiation Grant (previously known as the COIL grant). We also appreciate the support provided by Jared Stein from Instructure of Canvas.

Disclosure statement

No potential conflict of interest was reported by the authors.

Funding

This work was supported by the Penn State EdTech Network Initiation.

Notes on contributors

Hengtao Tang is a dual-title doctoral student in Learning, Design, & Technology and Comparative & International Education at the Pennsylvania State University, USA. His research interests include learning analytics, CSCL, STEM education, and knowledge structure.

Wanli Xing is an Assistant Professor in the Department of Educational Psychology and Leadership at Texas Tech University, USA, with a background in statistics, computer science and mathematical modeling. His research interests are educational data mining, learning analytics, and CSCL.

Bo Pei is a doctoral student in Instructional Technology at Texas Tech University, USA. His research interests are educational data mining, learning analytics, and CSCL.

ORCID

Hengtao Tang http://orcid.org/0000-0002-8846-7654

References

Anderson, A., Huttenlocher, D., Kleinberg, J., & Leskovec, J. (2014). Engaging with massive open online courses. In C.-W. Chung (Ed.), *Proceedings of the 23rd international conference on World*

Wide Web (pp. 687–698). New York, NY: Association for Computing Machinery. Retrieved from http://wwwconference.org/proceedings/www2014/forms/index.htm

Ball, G. H., & Hall, D. J. (1965). *ISODATA, a novel method of data analysis and pattern classification*. Menlo Park, CA: Stanford Research Institute.

Barak, M., Watted, A., & Haick, H. (2016). Motivation to learn in massive open online courses: Examining aspects of language and social engagement. *Computers & Education*, *94*, 49–60.

Barbera, E., Gros, B., & Kirschner, P. A. (2015). Paradox of time in research on educational technology. *Time & Society*, *24*, 96–108.

Bergner, Y., Kerr, D., & Pritchard, D. E. (2015). Methodological challenges in the analysis of MOOC data for exploring the relationship between discussion forum views and learning outcomes. In J. G. Boticario, & O. C. Santos (Eds.), *Proceedings of the 8th international conference on educational data mining* (pp. 234–241). Madrid, Spain: International Educational Data Mining Society. Retrieved from http://www.educationaldatamining.org/EDM2015/proceedings/full234-241.pdf

Brinton, C., Chiang, M., Jain, S., Lam, H., Liu, Z., & Wong, F. (2014). Learning about social learning in MOOCs: From statistical analysis to generative model. *IEEE Transactions on Learning Technologies*, *7*, 346–359.

Campbell, J., Gibbs, A. L., Najafi, H., & Severinski, C. (2014). A comparison of learner intent and behaviour in live and archived MOOCs. *The International Review of Research in Open and Distributed Learning*, *15*(5), 235–262.

Canal, L., Ghislandi, P., & Micciolo, R. (2015). Pattern of accesses over time in an online asynchronous forum and academic achievements. *British Journal of Educational Technology*, *46*, 619–628.

Cohen, J. (1960). A coefficient of agreement for nominal scales. *Educational and Psychological Measurement*, *20*, 37–46.

Csikszentmihalyi, M. (1975). *Beyond boredom and anxiety*. San Francisco, CA: Jossey-Bass.

Duda, R. O., & Hart, P. E. (1973). *Pattern classification and scene analysis*. New York, NY: John Wiley and Sons.

Engle, D., Mankoff, C., & Carbrey, J. (2015). Coursera's introductory human physiology course: Factors that characterize successful completion of a MOOC. *The International Review of Research in Open and Distributed Learning*, *16*(2), 46–68.

Genolini, C., & Falissard, B. (2010). KmL: K-means for longitudinal data. *Computational Statistics*, *25*, 317–328.

Glaser, B., & Strauss, A. (1967). *The discovery of grounded theory: Strategies for qualitative research*. Chicago, IL: Aldine.

Goggins, S., & Xing, W. (2016). Building models explaining student participation behavior in asynchronous online discussion. *Computers & Education*, *94*, 241–251.

Hampel, R., & Pleines, C. (2013). Fostering student interaction and engagement in a virtual learning environment: An investigation into activity design and implementation. *Calico Journal*, *30*, 342–370. Retrieved from http://www.jstor.org/stable/calicojournal.30.3.342

Huang, J., Dasgupta, A., Ghosh, A., Manning, J., & Sanders, M. (2014, March). Superposter behavior in MOOC forums. In *L@S '14 proceedings of the first ACM conference on learning @ scale conference* (pp. 117–126). New York, NY: Association for Computing Machinery. doi:10.1145/2556325.2566249

Kapur, M., Voiklis, J., & Kinzer, C. K. (2008). Sensitivities to early exchange in synchronous computer-supported collaborative learning (CSCL) groups. *Computers & Education*, *51*, 54–66.

Kizilcec, R. F., Piech, C., & Schneider, E. (2013). Deconstructing disengagement: Analyzing learner subpopulations in massive open online courses. In *Proceedings of the third international conference on learning analytics and knowledge* (pp. 170–179). New York, NY: Association for Computing Machinery. doi:10.1145/2460296.2460330

Kizilcec, R. F., & Schneider, E. (2015). Motivation as a lens to understand online learners: Toward data-driven design with the OLEI scale. *ACM Transactions on Computer-Human Interaction (TOCHI)*, *22*, 2.

Landis, J. R., & Koch, G. G. (1977). The measurement of observer agreement for categorical data. *Biometrics*, *33*, 159–174.

Liu, C. C., Chen, W. C., Lin, H. M., & Huang, Y. Y. (2017). A remix-oriented approach to promoting student engagement in a long-term participatory learning program. *Computers & Education*, *110*, 1–15.
Liu, C. C., Lin, C. C., Deng, K. Y., Wu, Y. T., & Tsai, C. C. (2014). Online knowledge sharing experience with creative commons. *Online Information Review*, *38*, 680–696.
Liu, C. C., Wang, P. C., & Tai, S. J. D. (2016). An analysis of student engagement patterns in language learning facilitated by Web 2.0 technologies. *ReCALL*, *28*, 104–122.
Mercer, N. (2008). The seeds of time: Why classroom dialogue needs a temporal analysis. *Journal of the Learning Sciences*, *17*, 33–59.
Molenaar, I. (2014). Advances in temporal analysis in learning and instruction. *Frontline Learning Research*, *2*(4), 15–24.
Phan, T., McNeil, S. G., & Robin, B. R. (2016). Students' patterns of engagement and course performance in a massive open online course. *Computers & Education*, *95*, 36–44.
Pingault, J. B., Côté, S. M., Vitaro, F., Falissard, B., Genolini, C., & Tremblay, R. E. (2014). The developmental course of childhood inattention symptoms uniquely predicts educational attainment: A 16-year longitudinal study. *Psychiatry Research*, *219*, 707–709.
Pingault, J. B., Tremblay, R. E., Vitaro, F., Carbonneau, R., Genolini, C., Falissard, B., & Côté, S. M. (2011). Childhood trajectories of inattention and hyperactivity and prediction of educational attainment in early adulthood: A 16-year longitudinal population-based study. *American Journal of Psychiatry*, *168*, 1164–1170.
Pintrich, P. R., & Schrauben, B. (1992). Students' motivational beliefs and their cognitive engagement in classroom academic tasks. In D. Schunk & J. Meece (Eds.), *Student perceptions in the classroom: Causes and consequences* (pp. 149–183). Hillsdale, NJ: Erlbaum.
Reimann, P. (2009). Time is precious: Variable-and event-centred approaches to process analysis in CSCL research. *International Journal of Computer-Supported Collaborative Learning*, *4*, 239–257.
Rosé, C. P., Carlson, R., Yang, D., Wen, M., Resnick, L., Goldman, P., & Sherer, J. (2014, March). Social factors that contribute to attrition in MOOCs. In *L@S '14, Proceedings of the first ACM conference on learning @ scale conference* (pp. 197–198). New York, NY: Association for Computing Machinery. doi:10.1145/2556325.2567879
Thomas, M. J. (2002). Learning within incoherent structures: The space of online discussion forums. *Journal of Computer Assisted Learning*, *18*, 351–366.
Tibshirani, R., & Walther, G. (2005). Cluster validation by prediction strength. *Journal of Computational and Graphical Statistics*, *14*, 511–528.
Vygotsky, L. S. (1978). *Mind in society: The development of higher mental processes*. Cambridge, MA: Harvard University Press.
Wise, A. F., Zhao, Y., Hausknecht, S. N., & Chui, M. M. (2013). Temporal considerations in analyzing and designing for online discussions in education: Examining duration, sequence, pace and salience. In E. Barbera & P. Reimann (Eds.), *Assessment and evaluation of time factors in online teaching and learning* (pp. 198–231). Hershey, PA: IGI Global.
Xie, K. U. I., DeBacker, T. K., & Ferguson, C. (2006). Extending the traditional classroom through online discussion: The role of student motivation. *Journal of Educational Computing Research*, *34*, 67–89.
Xing, W., Chen, X., Stein, J., & Marcinkowski, M. (2016). Temporal predication of dropouts in MOOCs: Reaching the low hanging fruit through stacking generalization. *Computers in Human Behavior*, *58*, 119–129.
Xing, W., Guo, R., Petakovic, E., & Goggins, S. (2015). Participation-based student final performance prediction model through interpretable genetic programming: Integrating learning analytics, educational data mining and theory. *Computers in Human Behavior*, *47*, 168–181.
Xiong, Y., Li, H., Kornhaber, M. L., Suen, H. K., Pursel, B., & Goins, D. D. (2015). Examining the relations among student motivation, engagement, and retention in a MOOC: A structural equation modeling approach. *Global Education Review*, *2*(3), 23–33. Retrieved from http://ger.mercy.edu/index.php/ger/index
Yang, D., Sinha, T., Adamson, D., & Rose, C. P. (2013, December). *Turn on, tune in, drop out: Anticipating student dropouts in massive open online courses*. Paper presented at the NIPS

data-driven education workshop, Lake Tahoe, NV. Retrieved from http://www.cs.cmu.edu/~diyiy/docs/nips13.pdf

Yousef, A. M. F., Chatti, M. A., Schroeder, U., & Wosnitza, M. (2014). What drives a successful MOOC? An empirical examination of criteria to assure design quality of MOOCs. In *Proceedings of 2014 IEEE, 14th international conference on advanced learning technologies* (pp. 44–48). Athens, Greece: IEEE. doi:10.1109/ICALT.2014.23

Zhang, J., Skryabin, M., & Song, X. (2016). Understanding the dynamics of MOOC discussion forums with simulation investigation for empirical network analysis (SIENA). *Distance Education*, *37*, 270–286.

Zhu, M., Bergner, Y., Zhang, Y., Baker, R., Wang, Y., & Paquette, L. (2016). Longitudinal engagement, performance, and social connectivity: A MOOC case study using exponential random graph models. In *LAK '16, Proceedings of the sixth international conference on learning analytics & knowledge* (pp. 223–230). New York, NY: Association for Computing Machinery. doi:10.1145/2883851.2883934