Article

Journal of Educational Computing Research
0(0) 1–22

DOI: 10.1177/0735633118784705
journals.sagepub.com/home/jec
SAGE

Time Really Matters: Understanding the Temporal Dimension of Online Learning Using Educational Data Mining

Hengtao Tang[1], Wanli Xing[2], and Bo Pei[2]

Abstract

Learning and participation are inseparable in online environments. To improve online learning, much effort has been devoted to encouraging online participation. However, previous research has investigated participation from a variable-based perspective, looking only for relationships between participation and other variables. Time can change and even reverse this relationship, however. Online learning is a cumulative process, but participation at several critical moments is more significant for learning than is participation at other points. To fully understand how learning unfolds over time, it is necessary to shift to a new perspective on learning. Adopting an event-based view on which the units of analysis are separate but interrelated learning events, this study investigates longitudinal patterns in online participation. Using data mining techniques for education, the study validated longitudinal patterns of participation as an accurate measure for differentiating learner performance. In addition, the first segment of this online learning experience was identified as the most critical moment in which educators should provide efficient interventions to help their learners maintain active participation. Additional design, pedagogical, and methodological implications for online teaching and facilitation practices are discussed at the end.

[1]Department of Learning and Performance Systems, Pennsylvania State University, University Park, PA, USA
[2]Department of Educational Psychology and Leadership, Texas Tech University, Lubbock, TX, USA

Corresponding Author:
Hengtao Tang, Pennsylvania State University, 314 Keller Building, University Park, PA 16802, USA.
Email: hengtao.tang@psu.edu

Keywords
temporal dimension, time sequence, longitudinal participation, longitudinal clustering, educational data mining, online learning

Because learning and participation are inseparable in online settings, patterns in online participation are often used to predict learner performance (Hrastinski, 2009). Empirical evidence also shows that active participation in online courses can improve learners' academic achievement (Cheng & Chau, 2016; Davies & Groff, 2005). For example, Cheng and Chau (2016) found that online learners who participated more frequently in course activities demonstrated increased achievement and reported higher levels of satisfaction. However, participation in an online course is not a constant variable: The causal relationship between participation and performance is not always as simple as "more participation yields better performance" (Canal, Ghislandi, & Micciolo, 2015; Tang, Xing, & Pei, in press). Time, as a fundamental dimension of learning, can alter and even reverse this relationship.

In the fields of education and human development, *time* is a complex concept with many established components (Barbera, Gros, & Kirschner, 2012, 2015; Vygotsky, 1978). Even though the effect of time has been explored in a wide variety of areas (e.g., the maturity and improvement of human development in Barbera et al., 2015), the temporal dimension of online learning has remained underexplored (Barbera et al., 2012, 2015; Knight, Wise, Chen, & Cheng, 2015). Previous studies have partially considered a few superficial components of time, including the length of time a construct remains in effect and the average pace at which learning proceeds (Knight et al., 2015), but the temporal dimension of learning also involves concerns regarding the effect of the time sequence (Knight et al., 2015, Mercer, 2008; Wise, Zhao, Hausknecht, & Chui, 2013). To better understand the temporal dimension of learning, Reimann (2009) proposes an event-based view of learning that conceives of learning as "a development and progressive improvement" instead of as "a punctual and static fact" (Barbera et al., 2015, p. 4). On this view, an *event* is the unit of analysis, and each event matters for online learning (Tang et al., in press). This view also holds that learning unfolds over time as events accumulate. Most importantly, however, the order of the events has a distinct influence on learning (Reimann, 2009).

The validity of the existing research findings on the significance of online participation is limited because it neither considers the sequence of events through which learning occurs nor identifies the moments most important to learning (Molenaar, 2014). On an event-based view, participation is seen as a series of particular events, each of which has a distinct influence on the learning process and on learning outcomes (Barbera et al., 2015; Mercer, 2008; Reimann, 2009). In line with this view, researchers speculate that the outcome of learning is determined by several key moments in the learning process (Wise et al., 2013).

If these key moments can be identified, educators could optimally facilitate learner performance by providing relevant interventions at these moments. To identify these moments, researchers recommend looking into the longitudinal pattern of online learner activity as it offers a more precise view of how learning unfolds over time (Barbera et al., 2015; Canal et al., 2015). The longitudinal pattern of online participation varies by time (Wise et al., 2013). However, in the methodological view, traditional methods of data collection (e.g., surveys and focus-group interviews) and analysis (e.g., correlation analysis) might not be able to reveal these subtle but significant details of learning (Xing, Chen, Stein, & Marcinkowski, 2016).

The emerging field of educational data mining (EDM) promises to advance our understanding of the temporal dimension of participation by enabling the investigation of longitudinal patterns in participation (Knight et al., 2015; Xing et al., 2016). EDM incorporates a number of computing techniques and enables researchers to develop, research, and apply "computerized methods to detect patterns in large collections of educational data that would otherwise be hard or impossible to analyze due to the enormous volume of data within which they exist" (Romero & Ventura, 2013, p. 12). Using EDM, researchers will likely be able to granularly predict learner performance using low-level trace data collected during learning and then use these predictions to automatically deliver relevant interventions (Xing et al., 2016). For example, the longitudinal k-means clustering algorithm (KmL) is an unsupervised EDM technique that allows participants to be clustered on the basis of their longitudinal traits and then analyze each cluster as a whole (Genolini & Falissard, 2010). Although it has been seldom used in education research, KmL has been widely used in medical and psychiatric research (Pingault et al., 2011, 2014). This algorithm outperforms traditional data collection and analysis methods because it addresses longitudinal patterns of constructs and also deals with large data sets (Tang et al., in press). For this reason, KmL and other EDM algorithms allow researchers to overcome methodological concerns and investigate the temporal dimension of online participation.

The purpose of this study is thus to apply EDM techniques (i.e., KmL) to uncover the temporal dimension of online participation and to identify key moments in an online course that demanded relevant and timely intervention. The study has three main goals. First, this study using KmL clustered learners with similar longitudinal participation patterns to identify the subtle differences among learners that traditional methods might not be able to reveal. Second, this study reexamined the relationship between participation and performance in online courses, taking into account the temporal dimension of online learning. Third, this study identified critical moments that significantly determined the engagement trajectories of the online learners. The results can help teachers to design and implement online courses that will engage learners.

Literature Review

Learning and Participation in Online Settings

Participation is an important predictor of online learner performance (Cheng & Chau, 2016; Davies & Groff, 2005). Mostly in an asynchronous format, online learning extends the opportunity of participation beyond the barriers of time and place (Xie, DeBacker, & Ferguson, 2006). In such courses, each learner can dedicate as much time as they need and can learn and reflect at their own pace. However, while this time-independent course format enriches interactions and increases learning gains (Meyer, 2003; Xie et al., 2006), students in asynchronous courses suffer from the absence of synchronism and inadequate available time (Barbera et al., 2015). Asynchronous courses force learners to wait unpredictably long to hear from their peers, and they force learners to repeatedly access the course (Dringus & Ellis, 2010). In addition, the forum posts and other activities common to asynchronous courses become increasingly overwhelming for online learners, both numerically and cognitively, and require them to devote additional time to participation (Peters & Hewitt, 2010). Ironically, this time-independent format leads to procrastination, which further causes that many learners in online courses participate infrequently or even drop out (Michinov, Brunot, Le Bohec, Juhel, & Delaval, 2011). As a result, the benefits of asynchronous courses accrue exclusively to those learners who actively participate (Hew & Cheung, 2008).

Because poor participation frequently decreases the quality of online courses (Hew & Cheung, 2008; Peters & Hewitt, 2010), it is critical to help online learners sustain active participation over time. Most online learners require efficient instructional support to participate consistently, and the timing of this support is especially important (Xing et al., 2016). For this reason, research into the temporal dimension of participation in online courses is essential, but it has been overlooked by much effort underway to boost learner participation. Existing research generally relies on traditional methods to intervene with certain variables for the purpose of increasing the numerical count of overall frequency of online participation patterns, but these methods yield rather limited implications on enabling temporal support for learners (Molenaar, 2014). For example, while Hew and Cheung (2008) confirm that peer-led facilitation can increase the summative frequency of learner participation in online courses, they tend to focus only on the effect of this intervention and overlook the search for the optimal moment to make this intervention available for learners. Without timely intervention, some learners are likely to have difficulty in maintaining their participation. Therefore, it is necessary to revisit the temporal dimension of online participation by examining the key moments to intervene with online learning. In this way, researchers are likely to figure out how to optimally support online learners.

Time and the Event-Based View of Learning

Although time is an integral dimension of human development and education (Vygotsky, 1978), education research often neglects an important element of time: sequence (Barbera et al., 2015; Reimann, 2009). Reimann (2009) argues that this situation is the result of misunderstandings regarding the unit of analysis in educational research. He explains that while some researchers have viewed learning as a static entity and attempted to find causal relationships between learning inputs—for example, *cognitive skill* and *learning style*—and learning outputs—for example, *grades* and *knowledge acquisition*—learning is better viewed as an evolutionary process of educational events. The first view is referred to as the *variable-centered* view whose unit of analysis is the variable, and latter view is referred to one as the *event-based* view in which the event is the unit of analysis.

Research that assumes the variable-centered view of learning typically investigates ''the extent to which individual learning skills (exogenous independent variable) can predict learning outcomes (dependent variable), dependent on more or less successful group communication (endogenous independent variable)'' (Reimann, 2009, p. 244). The variable-centered view's unit of analysis is the *variable*. Accordingly, its central focus is on causal relationships between independent and dependent variables. On this view, what matters for learning are the variances in the values of the variables between the starting and the ending point (Molenaar, 2014). The variable-centered view relies, however, on the fundamental assumption that causal relationships between independent and dependent variables are continuous and experience no interruptions (Reimann, Yacef, & Kay, 2011). This assumption is seldom or never met in online courses because online learners are not required to participate continuously and cease their participation at will. In online settings, in other words, causation between independent and dependent variables is not constant but varying due to the impact of time (Barbera et al., 2015; Molenaar, Chiu, Sleegers, & van Boxtel, 2011). For this reason, the variable-centered view may be unable to accurate describe how learning unfolds over time.

In contrast, Reimann (2009) espouses an event-based view that conceives of the learning process as ''a developmental event sequence, not a change in values of process variables'' (p. 247). For example, this view understands self-regulated learning as a time-dependent, cyclic sequence of events instead of viewing it as a static entity, as does the variable-centered view (Molenaar, 2014). On the event-based view, events are distinctly defined upon a central entity or *actor*. A central entity or *actor* can be a person, a group, or an idea (Reimann, 2009). For instance, we can view a learner as an *actor* and their individual participation patterns during a specific period as *events*. Using the *event* as the unit of analysis provides the event-based view noticeable advantages over the variable-based view (Barbera et al., 2015; Mercer, 2008; Molenaar, 2014; Reimann, 2009). First, every event in the course of learning is given its due consideration, not just the starting and ending points. The relationships between variables can change or reverse at certain points, and the corresponding events are still meaningful to the entire learning process. Second, because the event-based view understands learning as cumulative, the sequence of the events impacts how the learning progresses and what the learning achieves (Ritter, Nerb, Lehtinen, & O'Shea, 2007). Researchers who acknowledge this are more likely to identify the moments that matter most for learning and, as a result, to provide scaffolds in these moments that can produce better learning outcomes (Wise et al., 2013).

It is also worth noting that the findings of studies that have adopted the variable-based view may be inconsistent (Molenaar, 2014). Because they have overlooked temporally varied patterns of learning, previous studies have not been able to provide a holistic and precise overview of how learning unfolds over time. In contrast, researchers who espouse the event-based view respect the temporal dimension of learning and conduct in-depth investigations into longitudinal learning patterns to identify relationships that are dynamic over time (Barbera et al., 2015). Taking time sequence into account might alter the constant relationship between variables drawn from research holding a variable-based view of learning. Participation by online learners varies over time and is markedly influenced by several key moments in the process (Wise, Perera, Hsiao, Speer, & Marbouti, 2012; Wise et al., 2013). Therefore, we speculate that whether learner participation is related to performance when the temporal perspective of participation is considered still needs proven.

Research Questions

As a result of the rise of EDM techniques, researchers can now uncover the temporal dimensions of learning with greater granularity. This study uses the KmL algorithm to cluster online learners by their longitudinal patterns of online participation and then analyzes each cluster as a whole to answer the following three research questions:

1. What is the longitudinal trajectory of learner participation in an online learning environment?
2. What is the relationship between a learner's longitudinal participation pattern and their performance level?
3. What is the critical moment that determines online learner participation?

Methodology

Context and Participants

The study used a dataset that was a part of the JuxtaLearn Project at a Spanish university (Martín, Gértrudix, Urquiza-Fuentes, & Haya, 2015). This project was designed using a constructivist approach and enabled the participants to learn by doing (Martín et al., 2015). The goal of this course was to help students investigate the concept of *usability* by producing a video with a relevant theme. A total of 111 participants were involved: 82 were enrolled in media and communication, and 29 were enrolled in computer engineering. Participants from two domains served different roles in the collaboration to create the videos, computer engineering students as subject experts, and media and communication students as video producers. These students were at different campuses, so their sole medium of collaboration was an online educational social network: *ClipIt* (http://clipit.es; Martín et al., 2015).

Variables

Participation. During the course of the learning, the participants performed different activities that advanced their shared goal of producing a video. These activities were divided into four categories: *create* (e.g., creating a forum thread or adding a new PDF or PPT or video file), *annotate* (e.g., chatting via private messages), *delete* (e.g., deleting a file or blog entry), and *update* (e.g., updating a profile or icon or posting a comment to the group discussion thread). In this constructivist environment, all activities devoted to the course goals were counted as *participation* (Jonassen & Land, 2000).

The project lasted for 3 months: from September 2013 to December 2013. To determine the longitudinal pattern of learner participation, the researchers divided the project into several consecutive segments. Given the volume of this data set, the researchers decided that the duration of each segment would be 3 weeks. As a result, the entire learning experience was split into four segments: 09/20–10/10, 10/11–10/31, 11/1–11/21, and 11/22–12/13. In addition, given the sample size ($n = 111$), the researchers decided to group participants into clusters using their longitudinal patterns of online participation. In this way, the researchers were able to investigate each cluster as a whole to identify its unique patterns and to provide exclusive scaffolds (Genolini & Falissard, 2010).

Table 1 presents the basic descriptive statistics for the participation levels in the segments. From Segment 1 to Segment 3, learner participation decreased

Table 1. Descriptive Statistics of Learner Participation.

Segments	Mean	Range	*SD*
Segment 1	7.91	32	8.28
Segment 2	4.37	35	6.58
Segment 3	0.84	9	1.54
Segment 4	13.36	289	27.79

Table 2. Detailed Procedures of Data Analysis.

Process	Methods
Explore the longitudinal trajectory of learners' temporal participation	Longitudinal k-means clustering algorithm
Identify the relationship between temporal participation and performance	Multivariate analysis of variance; *T* test of the peer assessment
Recognize critical moments of temporal participation	Stepwise discriminant analysis

significantly; during Segment 3, participants were seldom involved in the educational experience. However, in Segment 4, course participation increased significantly, with a maximum value approaching 289. In addition, the standard deviation for course participation was largest in Segment 4 and was much larger in this segment than in any other segment. The data for one student were removed from the analysis because of that student's almost total inactivity.

Performance. Peer assessment was the primary means of assessing the participants' performance levels. A higher average peer rating indicated that the participant performed better. Four facilitators supervised the course, but they did not evaluate the participants' projects.

Data Processing and Analysis

Table 2 provides an overview of the methods of data analysis used in this study. The methods used were primarily longitudinal cluster analysis and statistical analysis. A more detailed description of the data-analysis procedures follows.

In the first step, an unsupervised clustering algorithm (longitudinal k-means clustering) was used to determine the optimal number of clusters (Milligan & Cooper, 1985). This study applied the Davies–Bouldin index (Davies & Bouldin, 1979). Previous studies have used the same hierarchical technique and also the same method for plotting the optimization parameter (e.g., Baker & Hubert, 1975; Saarela & Kärkkäinen, 2017). The latter two methods rely primarily on personally defined accounts of the largest change in the value (see Figure 1(a); the optimal number is 5), but the Davies–Bouldin index is computed using "the ratio of the sum of within-cluster scatter to between-cluster separation" (Maulik & Bandyopadhyay, 2002, p. 1651) with the minimum value of the ratio indicating the optimal number of clusters (see Figure 1(b); the optimal number is 2). In this way, the Davies–Bouldin index decreases the error caused by human interpretation.

In this study, the variables describing participation may have included multiple dimensions, so the data set was normalized to eliminate differences in scales (Xing et al., 2016). Figure 2 presents the results of longitudinal participation

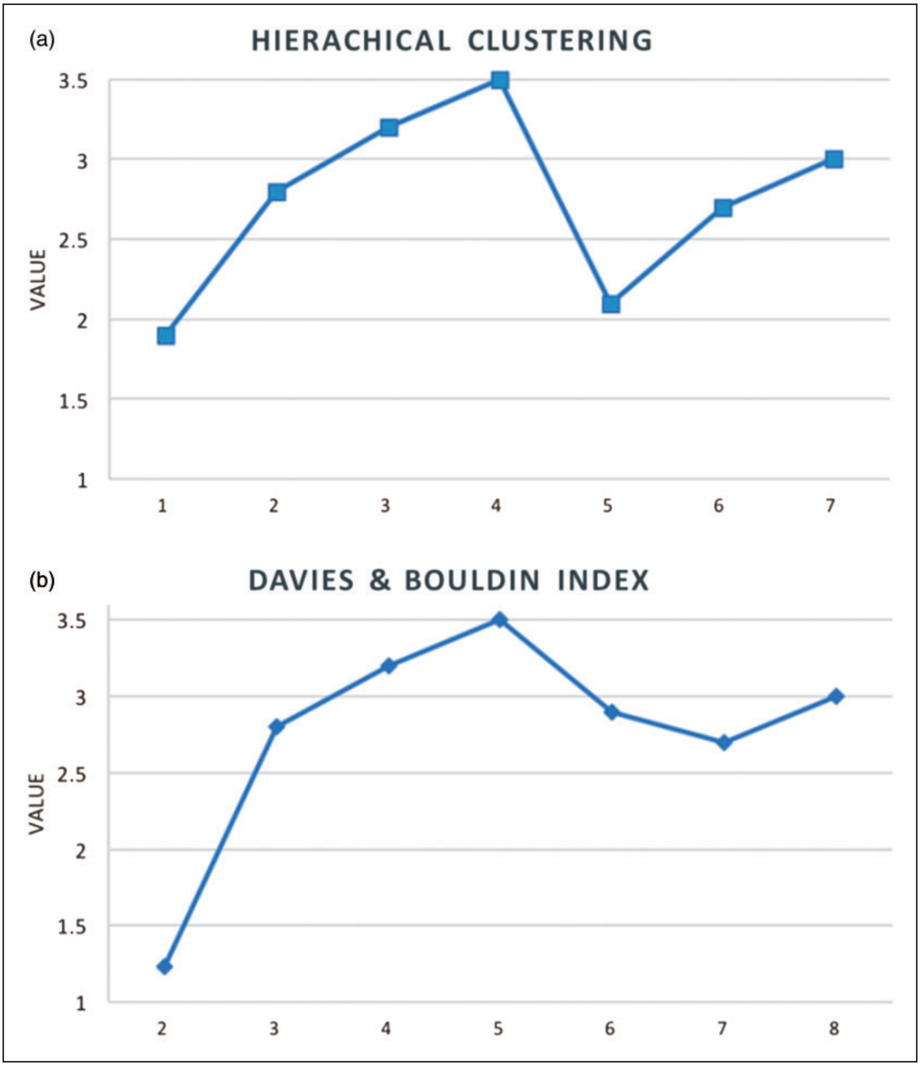

Figure 1. (a) Hierarchical clustering samples ($K = 5$). (b) Davies and Bouldin Index samples ($K = 2$). These figures illustrate that it is easier and more accurate to get the value of the optimal clustering number using Davies and Bouldin Index.

patterns after normalization. The X axis represents the time span of the four segments, and the Y axis indicates the normalized value of participation.

After the data set was normalized, the researchers used the longitudinal KmL to cluster the 111 participants into K groups using their longitudinal patterns of online participation (X). The KmL algorithm used in this study (see Figure 3)

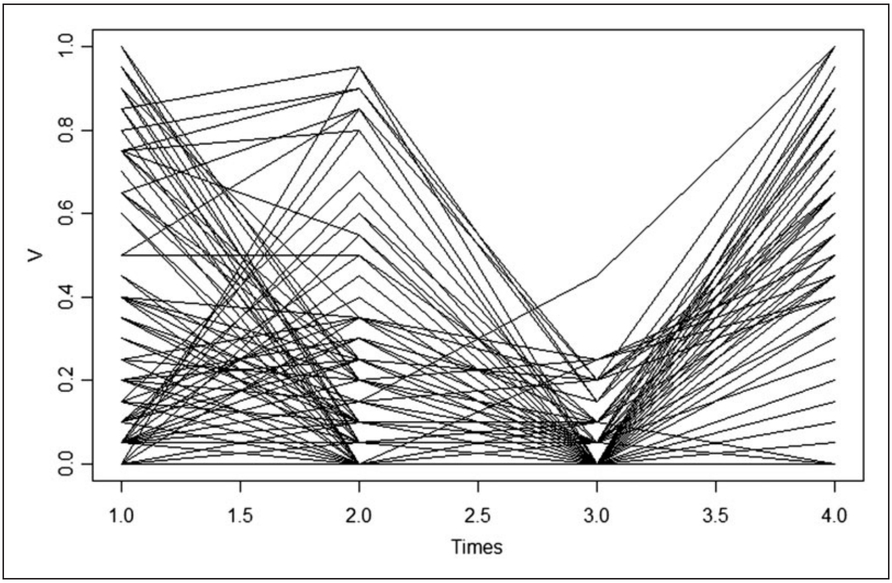

Figure 2. The longitudinal patterns of participation after the normalization between 0 and 1. This step is primarily to eliminate the difference in scales of variables.

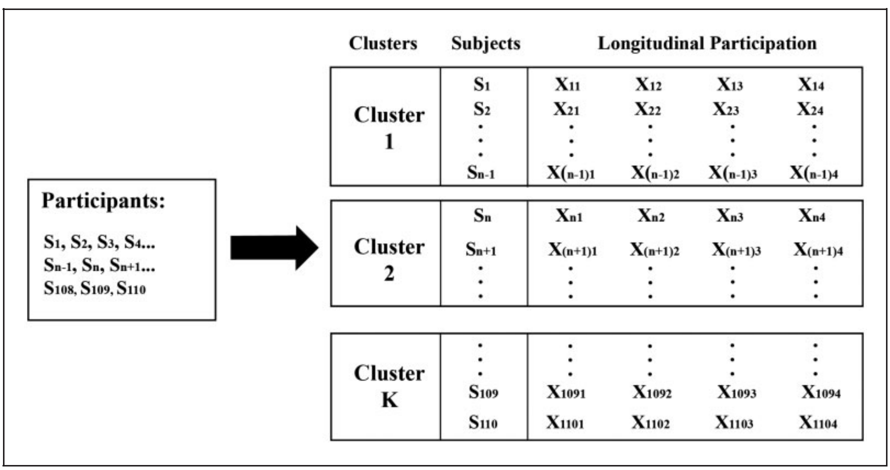

Figure 3. The illustration of the longitudinal k-means clustering algorithm (KmL). This figure illustrates how KmL clusters participants using their temporal participation patterns.

assessed the similarity between two participants using Euclidean distance (Genolini & Falissard, 2010). Euclidean distance, $Dist(X_n, X_m)$, was calculated using Equation (1), in which X_{nt} and X_{mt} represent the participation patterns associated with the two participants (n and m) at time point t. In addition, T represents the total number of segments (e.g., $T=4$). Participants n and m were more alike if their Euclidean distance was smaller and thus had a higher probability of being in the same cluster.

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

After clustering the participants into K groups, the researchers investigated whether these K groups demonstrated significant differences in performance and, if so, what caused this difference. Multivariate analysis of variance was used to determine whether there were significant differences in performance between clusters. Then, a two-sample T test was performed to determine whether the average performances of different clusters were statistically different. If a difference was significant, the researchers ran a stepwise linear discriminant analysis (SWLDA) to determine what caused the difference (e.g., when the critical moments were). The SWLDA algorithm is widely used to identify optimal predictor variables and is superior in its automaticity and efficient extraction of features (Krusienski, Sellers, McFarland, Vaughan, & Wolpaw, 2008; Siddiqi, Ali, Khan, Park, & Lee, 2015). Using the SWLDA, the researchers identified which segment was the critical moment during which to provide efficient interventions for the participants in these K groups.

Results

What Is the Longitudinal Trajectory of Learner Participation in an Online Learning Setting?

The results of the Davies–Bouldin index indicated that the optimal number of clusters was two. As indicated by the Davies–Bouldin index, the optimal number of clusters was found at the minimum point of the index (Davies & Bouldin, 1979). As shown in Figure 4, at the minimum point of the plot, the value was two.

The researchers used KmL to assign the participants to two clusters. Figure 5 presents the longitudinal patterns of participation for these clusters. A detailed description of the longitudinal variation of online participation is provided in Table 3.

Cluster A. About 66.4% of the participants were assigned to this cluster, which demonstrated a relatively lower longitudinal participation from the beginning

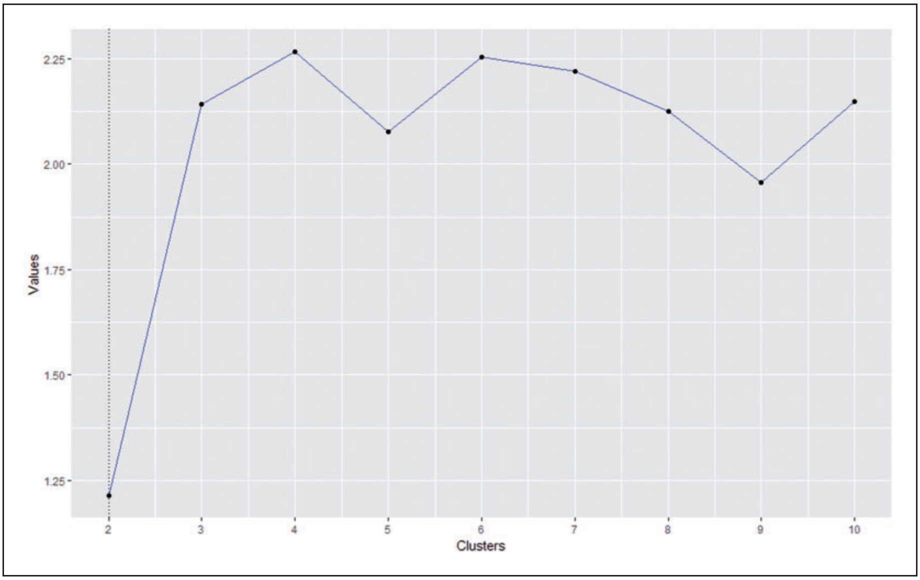

Figure 4. The Davies–Bouldin index determines the optimal number of clusters. This figure indicates that the optimal option is to assign participants into two clusters.

to the end. Specifically, the participants in Cluster A were much less active than their peers in Cluster B in the first segment, but they showed a slight increase in participation during the second segment. Cluster A had the lowest participation in the third segment. Participants in this cluster were barely involved during this period, as is indicated by their average value of participation, which was less than one. In fact, a large majority of the participants in Cluster A were not participating at all. However, the participants in Cluster A increased their participation significantly in the final segment, even to the highest level across the learning experience. In addition, during each of the segments, a subset of the participants in Cluster A were not at all involved in the experience.

Cluster B. About 33.6% of the participants were grouped into this cluster, which demonstrated relatively more active longitudinal participation throughout the four segments. The participants in Cluster B showed the highest participation during the first segment, and this value was the highest for either cluster throughout all of the segments. It is also worth noting that all of the participants in Cluster B were involved during the first segment (min = 8). The participation of Cluster B decreased during the second and the third segments, however. In the third segment, especially, the participation of Cluster B dropped sharply to its lowest level. Even here, however, the mean participation of Cluster B was higher than that of Cluster A. The participation of the participants in Cluster B

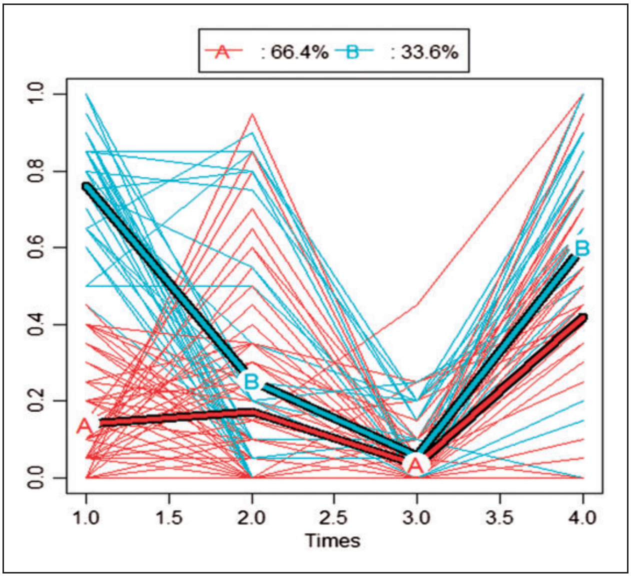

Figure 5. Two clusters of participants by longitudinal participation patterns. This figure shows the longitudinal trajectory of online participation of two clusters.

Table 3. Longitudinal Participation Values in Two Clusters.

	A					B				
Cluster	$N=73$					$N=37$				
Segment	Mean	Minimum	Maximum	Range	*SD*	Mean	Minimum	Maximum	Range	*SD*
1	2.79	0	9	9	2.53	15.24	8	20	12	2.98
2	3.41	0	19	19	4.69	5.05	0	18	18	5.94
3	0.68	0	9	9	1.53	1.14	0	5	5	1.53
4	8.41	0	20	20	5.52	12.05	0	20	20	4.35

returned to a relatively high level in the final segment, but these learners were still not as actively engaged as they had been in the first segment.

What Is the Relationship Between Learners' Longitudinal Participation Patterns and Their Performance Level?

The researchers then investigated whether the participants in the two clusters performed in significantly different ways. The participants' performance was

Table 4. MANOVA Results on Longitudinal Participation and Performance.

Effect	Value	F	Hypothesis df	Error df	Significance	Partial eta squared
Wilks' Lambda	0.16	140.96	4	105	.00*	0.84

Note. MANOVA = multivariate analysis of variance.
$^{*}p < .001$.

calculated by averaging the peer-assessment scores for the videos that their groups created. Each student was in the same group was given the same score. The researchers applied a multivariate analysis of variance to the clustered participants' performance. The results revealed that there was a statistically significant difference in performance based on the participants' longitudinal participation—$F(4, 105) = 140.96$, $p < .0005$, Wilk's $\Lambda = 0.16$, partial $\eta^2 = .84$ (see Table 4). Specifically, Cluster B performed better on average than did Cluster A (see Figure 6). A two-sample T test was conducted to determine whether this difference in average performance was significant. The results showed that the average performance of the participants in Cluster B was significantly higher than that of the participants in Cluster A ($t = 2.09$, $p < .05$).

What Is the Critical Moment That Determines Online Learner Participation?

The researchers identified the critical moments at which the participants required temporal interventions to encourage their active participation. The results of a stepwise linear discriminate analysis indicated that Segment 1 ($F = 527.74$, $p < .1$) had a much larger impact on learner participation than did any of the other segments (see Table 5). Segment 2 ($F = 3.84$, $p < .1$) and Segment 4 ($F = 2.91$, $p < .1$) also contributed to the difference in longitudinal participation between the two clusters, but their respective influences were lower than that of Segment 1 (see Table 5). Segment 3 was ranked last among the four segments. In short, Segment 1 was the most critical moment for differentiating students' longitudinal participation patterns, followed by Segment 2 and Segment 4. Figure 7(a) and (b) provides a visual explanation of the differences among the four segments.

Discussion

This study explored the temporal dimension of online learner participation patterns using EDM techniques. Specifically, it used the KmL algorithm to assign participants to two clusters based on their longitudinal pattern of online

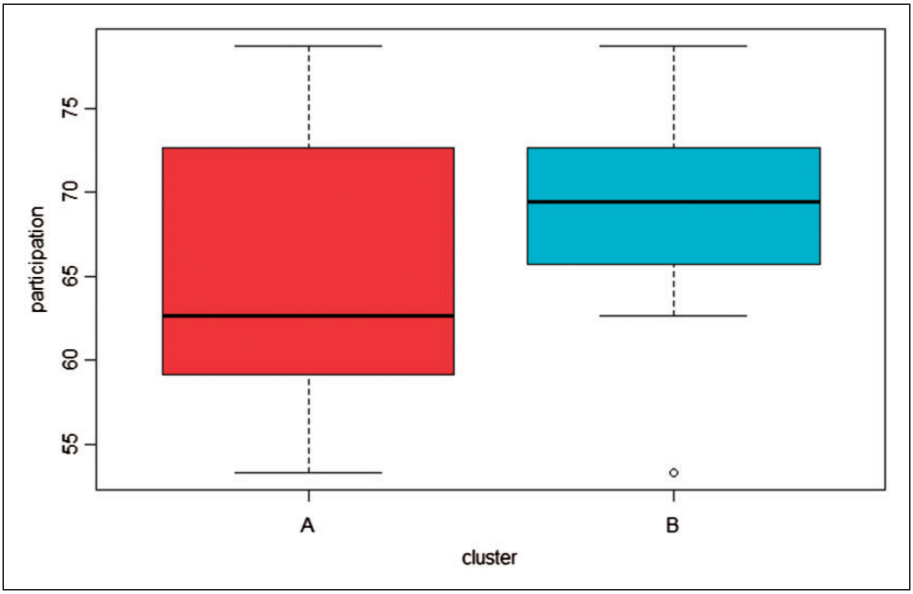

Figure 6. The boxplot of the difference in mean performance between Clusters A and B. This figure indicates that Cluster B performed better than Cluster A.

Table 5. Results of the Stepwise Linear Discriminant Analysis.

Steps	Segments	Wilks' Lambda	*F* difference	*p* difference
1	Segment 1	0.17	527.74	.00*
2	Segment 2	0.16	3.84	.05*
3	Segment 4	0.16	2.91	.09*
4	Segment 3	0.16	1.76	.19

*$p < .1$.

participation and validated this measure by establishing that it accurately differentiated learner performance. Throughout the learning experience, the participants in Cluster B were more active than were their peers in Cluster A. This result revealed that the participants with more active longitudinal participation generally outperformed their peers. This finding confirmed the importance of participation to online learner performance (Cheng & Chau, 2016; Davies & Groff, 2005; Hrastinski, 2009). In addition, this study advanced the event-based view of learning by considering each event in the longitudinal trajectory

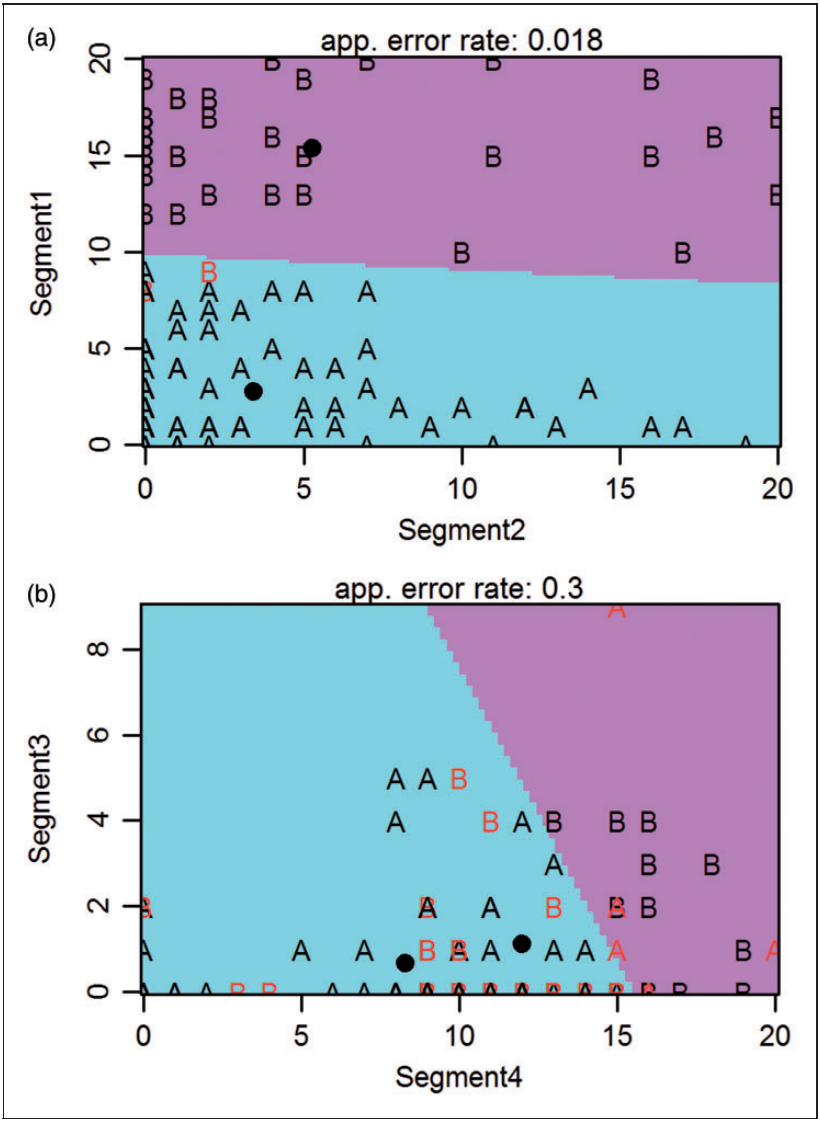

Figure 7. (a) and (b) The visualization of the temporal dimension of participation for two clusters in four segments. These two figures confirm Segment 1 plays the primary role in differentiating the longitudinal pattern of learner participation between two clusters.

of online participation, and it gave support to the idea that each event mattered in the longitudinal trajectory of learning (Molenaar, 2014; Reimann, 2009). Helping online learners to maintain longitudinal patterns of active participation improved their performance to a greater degree than does increasing the summative frequency of overall participation patterns.

Furthermore, Segment 1 was identified as the most important moment in differentiating the longitudinal patterns of learner participation, although Segments 2 and 4 also influenced the difference between the two clusters. During Segment 1, the learner participation of Cluster B was higher than that of either cluster in any of the other segments. Although the involvement of Cluster B decreased over the subsequent segments, Cluster B had a higher average participation than did Cluster A across all of the segments and outperformed Cluster A in this online learning experience.

These results suggest that the initial period of engagement during online learning might have a *downstream* effect on learners' subsequent participation and performance. In agreement with the event-based view of learning, we view learning as a cumulative process in which the temporal sequence of the events affects learning and its outcomes (Reimann, 2009). We also speculate that the learner engagement developed during the first segment may have been more effective than that developed during the other segments. Although the learners' participation fluctuated across the subsequent segments, their learning engagement was longitudinally influenced by its initial status: Higher initial engagement yielded a longitudinal pattern of relatively more active online participation (Cluster B was more engaged on average than was Cluster A).

In addition, the participation of both clusters increased remarkably in the final segment. This increase may have resulted from a desire on the part of the participants to complete the project and earn a satisfactory grade on the final assessment, but it could also provide an alternative viewpoint from which to understanding the drawbacks of time-independent online learning. Although time-independent online learning allows participants to regulate their participation at will (Xie et al., 2006), it also allows them to procrastinate (Michinov et al., 2011). This may be why both clusters produced their lowest average participation levels during Segment 3. The participants with inferior time-management and self-regulation skills may have waited until the last moment to dedicate an amount of effort sufficient to complete the video project. By taking this result into consideration and improving their course designs, educators can encourage their students to maintain longitudinal trajectories of active participation.

This study also provided educators with three empirical recommendations on efficiently engaging learners in online learning environments. First, online educators should grade learners' participation by grading their longitudinal trajectories of participation. While most online instructors allocate 10% to 20% of the final grade to participation, numerical convergence (a common way to grade learner participation) may not reveal subtle temporal changes in learner participation (Cheng & Chau, 2016; Davies & Graff, 2005). To solve this problem, this study offers an alternative method for accurately grading the participation of online learners. Second, to ensure that their online learners remain actively engaged, educators should provide facilitation or scaffolding in early stages. Empirical research recommends that educators diversify ongoing interactions between students (Macfadyen & Dawson, 2010) and that they increase their presence in the early period of a course (Ma, Han, Yang, & Cheng, 2015). Third, to prevent procrastination, which inhibits the development of longitudinal trajectories of active participation (Michinov et al., 2011), educators could create milestones for their courses and divide the final assessment among the resulting segments (Ariely & Wertenbroch, 2002). Another way to encourage active participation is to provide regular, formative feedback (Doherty, 2006).

In addition, this study supports the conclusion that EDM reveals the nature of learning more thoroughly than do traditional methods (Knight et al., 2015; Xing et al., 2016). The KmL algorithm is a promising EDM technique for education research for a number of reasons. First, because it yields more granular results and can better process large volumes of data, researchers can use the KmL algorithm to produce more trustworthy findings. Moreover, because EDM can detect low-level features—for example, the frequency with which someone visits a forum and the number of pages that person accesses—it can be used to build models that predict dropout from large online courses. Such models could allow intelligent agents to identify the dropout probability for each student and to automatically provide each student appropriate, predefined interventions (Xing & Du, in press). The automaticity of EDM gives it the potential to boost the participation of online learners. EDM has precipitated a growing interest in the temporal aspects of learning (Knight et al., 2015), and advanced computing methodologies and models are needed to perpetuate this momentum. While KmL is a relatively new EDM technique, it has been widely applied in other disciplines (Genolini & Falissard, 2010). Its potential in investigating the temporal dimension of learning could inspire multidisciplinary efforts to produce methodological breakthroughs in the field of EDM.

This study had several limitations. First, it focused solely on behavioral patterns in participation and did not examine the process of knowledge acquisition. Second, every participant in a given group received the same score, regardless of their individual contribution. Third, because this online experience included only a peer assessment and not a formal assessment, the score may have been less reliable. Fourth, because this study did not employ a formal assessment, it may have underestimated the influence of motivational factors on online learner behaviors. To avoid these limitations, future studies could obtain larger data sets from formal online courses. Moreover, future studies into online collaborative-learning environments could touch upon the temporal aspect of the process of knowledge construction.

Conclusion

Since learning and participation are inseparable in online learning environments, much effort has been devoted to increasing the participation of online learners (Hrastinski, 2009). However, many of these endeavors have overlooked the temporality of online learning. To fill this gap in the research, this study used EDM techniques to investigate online learners' longitudinal trajectories of participation and to identify critical moments in which educators can encourage active participation. Longitudinal participation patterns were validated as accurate measures for differentiating learner performance in online environments; learners who participate more actively are more likely to excel in online learning.

This study also found that the first segment was the most critical to learner performance. Accordingly, we suggest that educators provide efficient scaffolding in the early stages of their online learning experiences. Educators might also consider establishing milestones to help their online learners overcome the temptation to procrastinate (Michinov et al., 2011). As the results of this study showed, even though they were expected to remain actively involved throughout the learning experience, the online learners tended to wait until the end of the experience to significantly invest in their final project. This study had several limitations, including the design of the learning environment and the absence of a test for individual learning gains. Future studies could also select larger data sets and investigate gains in the construction of knowledge produced via online environments.

Acknowledgments

The research was supported by the Penn State EdTech Network Initiation Grant (previously known as the COIL grant). We also appreciate the dataset access support provided by Estefanía Martín, Manuel Gértrudix, Jaime Urquiza-Fuentes and Pablo A. Haya from the Universidad Rey Juan Carlos in Spain and also the support of Okan Arslan in preprocessing the data.

Declaration of Conflicting Interests

The authors declared no potential conflicts of interest with respect to the research, authorship, and/or publication of this article.

Funding

The authors disclosed receipt of the following financial support for the research, authorship, and/or publication of this article: This study was supported by the Penn State EdTech Network Initiation Grant.

ORCID iD

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

References

Ariely, D., & Wertenbroch, K. (2002). Procrastination, deadlines, and performance: Self-control by precommitment. *Psychological Science*, *13*(3), 219–224.

Baker, F. B., & Hubert, L. J. (1975). Measuring the power of hierarchical cluster analysis. *Journal of the American Statistical Association*, *70*(349), 31–38.

Barbera, E., Gros, B., & Kirschner, P. (2012). Temporal issues in e-learning research: A literature review. *British Journal of Educational Technology*, *43*(2), E53–E55. doi:10.1111/j.1467-8535.2011.01255.x

Barbera, E., Gros, B., & Kirschner, P. A. (2015). Paradox of time in research on educational technology. *Time & Society*, *24*(1), 96–108.

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. doi:10.1111/bjet.12158

Cheng, G., & Chau, J. (2016). Exploring the relationships between learning styles, online participation, learning achievement and course satisfaction: An empirical study of a blended learning course. *British Journal of Educational Technology*, *47*(2), 257–278.

Davies, D. L., & Bouldin, D. W. (1979). A cluster separation measure. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, *1*(2), 224–227. doi:10.1109/TPAMI.1979.4766909

Davies, J., & Graff, M. (2005). Performance in e-learning: Online participation and student grades. *British Journal of Educational Technology*, *36*(4), 657–663.

Doherty, W. (2006). An analysis of multiple factors affecting retention in web-based community college courses. *The Internet and Higher Education*, *9*, 245–255.

Dringus, L. P., & Ellis, T. (2010). Temporal transitions in participation flow in an asynchronous discussion forum. *Computers & Education*, *54*(2), 340–349.

Genolini, C., & Falissard, B. (2010). KmL: k-means for longitudinal data. *Computational Statistics*, *25*(2), 317–328.

Hew, K. F., & Cheung, W. S. (2008). Attracting student participation in asynchronous online discussions: A case study of peer facilitation. *Computers & Education*, *51*(3), 1111–1124.

Hrastinski, S. (2009). A theory of online learning as online participation. *Computers & Education*, *52*(1), 78–82.

Jonassen, D. H., & Land, S. M. (2000). *Theoretical foundations of learning environments*. Mahwah, NJ: Lawrence Erlbaum Associates.

Knight, S., Wise, A. F., Chen, B., & Cheng, B. H. (2015, March). It's about time: 4th international workshop on temporal analyses of learning data. In *Proceedings of the Fifth International Conference on Learning Analytics and Knowledge* (pp. 388–389). New York, NY: ACM.

Krusienski, D. J., Sellers, E. W., McFarland, D. J., Vaughan, T. M., & Wolpaw, J. R. (2008). Toward enhanced P300 speller performance. *Journal of Neuroscience Methods*, *167*(1), 15–21.

Ma, J., Han, X., Yang, J., & Cheng, J. (2015). Examining the necessary condition for engagement in an online learning environment based on learning analytics approach: The role of the instructor. *The Internet and Higher Education*, *24*, 26–34.

Macfadyen, L. P., & Dawson, S. (2010). Mining LMS data to develop an ''early warning system'' for educators: A proof of concept. *Computers & Education*, *54*(2), 588–599.

Martín, E., Gértrudix, M., Urquiza-Fuentes, J., & Haya, P. A. (2015). Student activity and profile datasets from an online video-based collaborative learning experience. *British Journal of Educational Technology*, *46*(5), 993–998.

Maulik, U., & Bandyopadhyay, S. (2002). Performance evaluation of some clustering algorithms and validity indices. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, *24*(12), 1650–1654.

Mercer, N. (2008). The seeds of time: Why classroom dialogue needs a temporal analysis. *The Journal of the Learning Sciences*, *17*(1), 33–59.

Meyer, K. (2003). Face-to-face versus threaded discussions: The role of time and higher-order thinking. *Journal for Asynchronous Learning Networks*, *7*(3), 55–65.

Michinov, N., Brunot, S., Le Bohec, O., Juhel, J., & Delaval, M. (2011). Procrastination, participation, and performance in online learning environments. *Computers & Education*, *56*(1), 243–252.

Milligan, G. W., & Cooper, M. C. (1985). An examination of procedures for determining the number of clusters in a data set. *Psychometrika*, *50*(2), 159–179.

Molenaar, I. (2014). Advances in temporal analysis in learning and instruction. *Frontline Learning Research*, *2*(4), 15–24.

Molenaar, I., Chiu, M. M., Sleegers, P., & van Boxtel, C. (2011). Scaffolding of small groups' metacognitive activities with an avatar. *International Journal of Computer-Supported Collaborative Learning*, *6*(4), 601–624.

Peters, V. L., & Hewitt, J. (2010). An investigation of student practices in asynchronous computer conferencing courses. *Computers & Education*, *54*(4), 951–961.

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*(3), 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*(11), 1164–1170.

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*(3), 239–257.

Reimann, P., Yacef, K., & Kay, J. (2011). Analyzing collaborative interactions with data mining methods for the benefit of learning. In S. Puntambekar, G. Erkens & C. E. Hmelo-Silver (Eds.), *Analyzing interactions in CSCL: Methods, approaches and issues* (pp. 161–185). New York, NY: Springer.

Ritter, F. E., Nerb, J., Lehtinen, E., & O'Shea, T. (Eds.). (2007). *In order to learn: How the sequences of topics influences learning*. Oxford, England: Oxford University Press.

Romero, C., & Ventura, S. (2013). Data mining in education. *Wiley Interdisciplinary Reviews: Data Mining and Knowledge Discovery*, *3*(1), 12–27.

Saarela, M., & Kärkkäinen, T. (2017). Knowledge discovery from the programme for international student assessment. In A. Peña-Ayala (Ed.), *Learning analytics: Fundaments, applications, and trends* (pp. 229–267). New York, NY: Springer International Publishing.

Siddiqi, M. H., Ali, R., Khan, A. M., Park, Y. T., & Lee, S. (2015). Human facial expression recognition using stepwise linear discriminant analysis and hidden conditional random fields. *IEEE Transactions on Image Processing*, *24*(4), 1386–1398.

Tang, H., Xing, W., & Pei, B. (in press). Exploring the temporal dimension of forum participation in MOOCs. *Distance Education*, https://doi.org/10.1080/01587919.2018.1476841

Vygotsky, L. S. (1978). *Mind in society: The development of higher psychological processes*. London, England: Harvard University Press.

Wise, A. F., Perera, N., Hsiao, Y. T., Speer, J., & Marbouti, F. (2012). Microanalytic case studies of individual participation patterns in an asynchronous online discussion in an undergraduate blended course. *The Internet and Higher Education*, *15*(2), 108–117.

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., DeBacker, T. K., & Ferguson, C. (2006). Extending the traditional classroom through online discussion: The role of student motivation. *Journal of Educational Computing Research*, *34*(1), 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. & Du, D. (in press). Dropout prediction in MOOCs: Using deep learning for personalized intervention. *Journal of Educational Computing Research*, https://doi.org/10.1177/0735633118757015.

Author Biographies

Hengtao Tang is a dual-title doctoral student in Learning, Design, and Technology and Comparative and International Education at the Pennsylvania State University. His research interests include online education, learning analytics, learning science, and science, technology, engineering, and mathematics education.

Wanli Xing is an assistant professor in Instructional Technology at Texas Tech University, USA with background in learning sciences, statistics, computer science, and mathematical modeling. His research interests are educational data mining, learning analytics, and computer-supported collaborative learning.

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