Internet and Higher Education 29 (2016) 1–11

Contents lists available at ScienceDirect

Internet and Higher Education

Clustering blended learning courses by online behavior data: A case study in a Korean higher education institute☆

Yeonjeong Park [a,1], Ji Hyun Yu [b,1], Il-Hyun Jo [c,*]

[a] Ewha Education Research Institute, Ewha Womans University, South Korea
[b] Faculty of Learning Sciences and Technologies, Virginia Tech., USA
[c] Department of Educational Technology, Ewha Womans University, South Korea

ARTICLE INFO

Article history:
Received 25 February 2015
Received in revised form 31 October 2015
Accepted 2 November 2015
Available online 3 November 2015

Keywords:
Blended learning
Higher education
Academic analytics
Educational data mining
Latent class analysis

ABSTRACT

Blended learning (BL) is recognized as one of the major trends in higher education today. To identify how BL has been actually adopted, this study employed a data-driven approach instead of model-driven methods. Latent Class Analysis method as a clustering approach of educational data mining was employed to extract common activity features of 612 courses in a large private university located in South Korea by using online behavior data tracked from Learning Management System and institution's course database. Four unique subtypes were identified. Approximately 50% of the courses manifested *inactive* utilization of LMS or *immature* stage of blended learning implementation, which is labeled as Type I. Other subtypes included Type C — Communication or Collaboration (24.3%), Type D — Delivery or Discussion (18.0%), and Type S — Sharing or Submission (7.2%). We discussed the implications of BL based on data-driven decisions to provide strategic institutional initiatives.

1. Introduction

Blended learning is a newly accepted approach in higher education (Graham, Woodfield, & Harrison, 2013; Norberg, Dziuban, & Moskal, 2011; Ross & Gage, 2006). While early studies (Beisser & Steinbronn, 2002; Cameron, 2003; Carroll & Hsu, 2003; Tuckman, 2002) introduced blended learning as a mere combination of classroom activities and online activities, many institutions of higher education are seriously considering as a way to transform traditional pedagogy for both on campus and distance education modes and thus maximize student learning and success (Garrison & Vaughan, 2013). According to the EDUCAUSE report entitled *Undergraduate Students and Information Technology* (Dahlstrom, Walker, & Dziuban, 2013), nearly 80% of U.S. college students have taken at least a blended learning course, indicating the desire for mixed modalities to deepen their learning engagement. To meet these expectations, this report proposes that institutions need to "evaluate whether current services and support practices are adapted for blended modalities" (p. 36). More importantly, today's students appreciate the value of learning management system (LMS) as digital learning environments in which "mobile-friendly, highly personalized, and engaging learning experiences" are enabled (p. 23). For instance, students can be provided with real-time feedback on their learning progress through personalized dashboards in the LMS. These features also can help not only instructors in improving their pedagogical practices to increase student engagement, but also the whole institution in supporting students' success in individual courses, in their college experience, and in their educational goals. However, very few institutions have a unified approach to blended learning in which students can have transformational learning experiences; and this may relate to the fact that many institutions have rarely made full use of the LMS as an enterprise system for institutional analytics. Rather, the ways to design blended learning course activities and employ LMS features are normally adopted by individual faculty, and, in most cases, instructors simply transfer course content into the LMS only using basic features such as posting their syllabuses and uploading lecture notes (Dahlstrom, Brooks, & Bichsel, 2014; Graham et al., 2013). This has led to lack of understanding about: (a) which instructional interventions have been *actually* adopted for blended learning within an institution and (b) how the institution leverage the LMS features to promote institutional approach and support mechanism for blended learning (Porter, Graham, Spring, & Welch, 2014).

In response to this concern, a great deal of research efforts have been made to advance blended learning (Cooner, 2010; Donnelly, 2010; Dziuban, Hartman, & Moskal, 2004; Graham, 2006; Hofmann, 2006; Poon, 2013; Shank, 2011; Vaughan, 2010). The special issue of *The*

☆ This research was supported by the National Research Foundation of Korea grant funded by the Ministry of Education, Science, and Technology (NRF no. 2013S1A5A2A0304410).

* Corresponding author at: Department of Educational Technology, Ewha Womans University, 52 Ewhayeodae-gil, Seodaemun-gu, Seoul 03760, South Korea.
E-mail address: ijo@ewha.ac.kr (I.-H. Jo).
[1] Co-first authors.

http://dx.doi.org/10.1016/j.iheduc.2015.11.001
1096-7516/

Internet and Higher Education in 2013 was dedicated to the topic of *Blended Learning Policy and Implementation Cases*, indicating possible reasons why blended learning has not been successfully implemented with strategic plans despite its many innate advantages. The primary method of these studies involved self-reported surveys and interviews with students, faculty, and administrative officials. Such traditional reliance on perceptions and anecdotes has hindered a more systematic and simultaneous evaluation of the current services for blended learning experiences. Accordingly, in this study, we aim to improve upon previous research, making *empirical* and *methodological* progress, using learning analytics techniques that are based on the use of large-scale data from the LMS in one institution of higher education. This data-driven approach might account for whether and how the institution invest in blended learning to support student success and how they can leverage the LMS features to make their educational services more affordable and effective.

2. Analytics in blended learning courses

2.1. Definition and categorization of blended learning

A literature has defined blended learning as a combination of face-to-face and online learning instruction with the aim of complementing each other (Graham et al., 2013). While the definition is clear and simplistic, its implementation is complex and rather challenging since virtually limitless designs are possible depending on how much or how little *online instruction* is inherent in blended learning (Garrison & Kanuka, 2004). Diverse instructional models and best practices of blended learning have been reported from simple use of computer or online mediated technologies to full usages of them for a complete course.

Recent thematic analysis on blended learning conducted by Halverson, Graham, Spring, Drysdale, and Henrie (2014) indicated that, during the recent decade, 41% of blended learning research had questions about instructional design including models, strategies, best practices, implementation, and environment & course structure. As an example, in regard to identifying the types of blending, Singh (2003) introduced the following five dimensions: simple blending of offline and online (1st level), blending self-paced and live collaborative learning (2nd level), blending structured and unstructured learning (3rd level), blending custom content with off-the-shelf content (4th level), and blending learning, practice, and performance support (5th level). From a broader perspective, Margulieux, Bujak, McCracken, and Majerich (n.d) distinguished several terms such as hybrid, blended, flipped, and inverted into a framework based on two dimensions: 1) information transmission vs. praxis, and 2) delivery via instructor vs. delivery via technology.

The learning experience taxonomy that they adopted suggested the following four major types of learning: face-to-face mixed (e.g., course with lab), lecture hybrid (e.g., part F2F, part online lecture), practice hybrid (e.g., part F2F, part online praxis), and online mixed (e.g., MOOC). Recently, flipped classroom gained attention as a new pedagogical method. It employs asynchronous video lectures and problem-solving practices as students' homework with active and diverse group-based activities in the classroom. These pedagogical approaches represent not only a combination of instructional methods online and offline, but also a combination of learning theories such as problem-based approaches based on constructive ideology vs. traditional lectures derived from direct instruction method based on behaviorist principles (Bishop & Verleger, 2013).

In summation, the current definitions and taxonomy of blended learning include a broad spectrum both in the *delivery* modalities between offline and online and the *pedagogies* between instructor-led and student-centered approaches. These two spectrums are marked in x-axis and y-axis respectively in Fig. 1. Four possible combinations have coexisted so far in the representative 4 types including 1) mostly face-to-face class with substantial online activities, 2) mostly online class with student offline group meeting, 3) mostly face-to-face lecture with online resources, and 4) mostly online lecture with optional face-to-face meeting. In this study, we do not stick to the one level or pedagogical taxonomy of blended learning dimensions or terms, rather we allow a variety of possibilities illustrated in Fig. 1 and attempt to investigate the clusters of blended learning courses in a higher education institute in Korea.

2.2. Adoption framework of blended learning

While the conceptual framework summarized in Fig. 1 is useful to capture the phenomenon of blended learning, it is limited to presenting the adoption level of online instruction. Graham et al. (2013) illustrated blended learning adoption spectrum with awareness and exploration (level 1), adoption and early implementation (level 2), and mature implementation and growth (level 3). Francis and Raftery (2005) also introduced three e-learning modes of engagement in the landscape of blended learning: (1) baseline course administration and learner support; (2) blended learning leading to significant enhancements to learning and teaching process, and (3) all of two modes to the level of personalized instruction through diverse online courses and modules. With regard to these adoption phase-based categorization founded upon *Diffusion of Innovation* theory (Rogers, 2010), the implications suggest that the university administrators need to be aware of such phases in the aspect of *strategy*, *structure*, and *support* so that they can make a transition from the lower to upper level for the high quality teaching and learning environment (Graham et al., 2013).

2.3. Institutional data-driven approach

Previous studies mentioned above have classified blended learning in diverse frameworks and approaches. Although inconsistent terminological usages and disagreement on identifying blended learning between the replacement of F2F time by online instruction and the supplementary role of online instrument might hinder further discussions, it is noticeable that most approaches have identified the types of blending. The level of adoption tends to be *theory-driven* rather than data-driven. They were focused on course-level analysis. Evidence-based empirical researches and verification of theoretical frameworks are certainly necessary. Such needs have also been addressed by assertions that there is a dearth of research guiding institutions to prepare strategic adoption of blended learning throughout institution-wide investigations with 6 cases (Graham et al., 2013) or 11 cases (Porter et al., 2014) selected from U.S. institutions. As a similar approach, Nichols (2008) investigated e-learning diffusions with 14 educational institutions including 8 universities and found considerable diversities, ranging from some institutions where e-learning was absorbed into daily activity to other institutions where e-learning was treated as an *external* or ad hoc activity by some enthusiastic group of people.

Few research studies have focused on technology adoption status in a single higher education institution with the attempt to investigate how different LMSs are used in practice. Carvalho, Areal, and Silva (2011) examined the extent and depth of the use of two different LMSs by investigating college students' perceptions and experiences in a university located in Portugal. Their results indicated a basic utilization level including dominant perceptions of Mode 1, with little instances of Mode 2 indicated by the aforementioned Francis and Raftery (2005)'s framework. They asserted that “the fact that [an online] learning platform is available, or that it is used extensively, does not necessarily mean they are used to similar levels of engagement” (p. 825). Blin and Munro (2008) also highlighted that there was little evidence of significant impact on technology disruption. They investigated an institution-wide deployment of Virtual Learning Environment (VLE) by using both numerical data from the VLE database and survey of staff users. In their

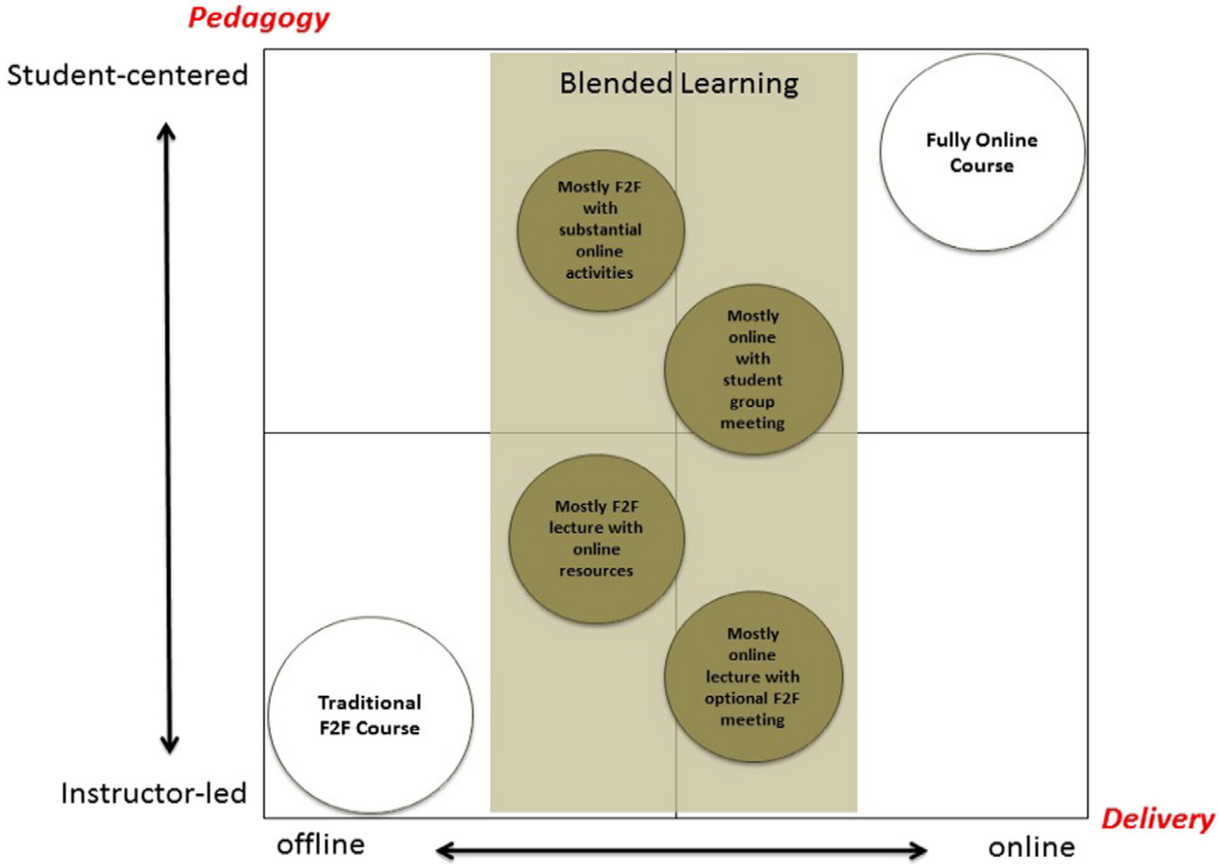


Fig. 1. Spectrum: Blended learning definitions.

institution, VLE was mainly used for administrative purposes to disseminate resources or information rather than used for the activities that demanded collaboration or reflection such as wikis or journals.

In spite of the meaningfulness of the evidence-based systematic investigation that reveal the status of blended learning or technology integration and the usefulness of adoption-based conceptual framework, the difficulties of investigation or dependence of subjective perceptions still suggest the use of advance technologies with available data tracked from the existing system.

2.4. Academic analytics and educational data mining

Research on institutional BL adoption, specifically the implementation of online intervention, can be leveraged by using data extracted from academic system. When compared to human observation and self-reported survey methods, the use of log data from the system such as LMS or VLE is more extensive in the number of participants, more comprehensive in the number of sessions, and more exquisite in the level of details (Mostow et al., 2005). The term "Academic Analytics" was first coined to describe the intersection for technology, information, management culture, and the application of information to manage the academic enterprise. It was derived from business intelligence (Goldstein & Katz, 2005, p. 2). Campbell, DeBlois, and Oblinger (2007) introduced Academic Analytics as a new tool to respond to increased concerns for accountability in higher education and to develop actionable intelligence to improve student success and learning environment. Most of its early applications were to predict student's academic success or to identify at-risk students for special guidance from their faculty and advisors. The Early Warning System such as Course Signal of Purdue University is a representative example that Academic Analytics is well applied. According to Arnold (2010), Signals work by mining data from multiple data sources such as Course Management System (CMS), Students Information System (SIS), and the grade book. Data cleansing, transformation, and algorithm generate a risk level of each student with green, yellow, or red indicator. She reported that administrators could consider academic analytics as "a scalable solution to support student success, familiarize students with campus help resources and improve the fail/withdraw rates of large-enrollment, low-interaction courses".

Baepler and Murdoch (2010) highlighted that, as in Purdue's Signal case, academic analytics, educational data mining technique, and CMS audits could be useful to guide course redesign and to implement evidence-based decisions in higher education. They especially distinguished academic analytics from educational data mining. The former is "a scientific, hypothesis-driven approach using a particular dataset to solve a practical academic problem, such as increasing student retention levels … [while the latter is] speculative prospecting for riches …to shift through data for implicit affinities and hidden patterns without a preconceived hypothesis" (p. 4). On the other hand, Long and Siemens (2011) compared learning and academic analytics. They stated that "academic analytics reflects the role of data analysis at an institutional level, whereas learning analytics centers on the learning process which includes analyzing the relationship between learner, content, institution, and educator" (p. 34).

3. Research purpose and questions

Following the literature review, we hypothesized that the instructionally distinct subgroups of blended learning courses could be identified based on course characteristics and requirements as well as associated student learning behaviors. Goldstein and Katz (2005) indicated that most institutions were staying at either stage 1 (extraction and reporting of transaction-level data) or stage 2 (operational reporting), while few institutions utilized academic analytics for stage

3 (what-if decision support) or higher stages such as stages 4 (predictive modeling and simulation) and stage 5 (automatic triggers of business process such as alerts). This study is a steppingstone to move stage 1 or 2 toward 3 or higher because sophisticated prediction model and interventions (e.g., development of successful alert system) may require precise diagnosis of the current status based on the understanding of diverse evolution patterns of blended learning courses. Thus, we applied a latent class analysis (LCA) which is a method of categorizing individual cases from a heterogeneous population into smaller, relatively homogenous subgroups. The following research questions guided the design of this study: (1) What are the demographic and instructional characteristics of blended learning courses? and (2) What usage patterns and clusters emerge across blended learning courses, when mining students' participation data related to online learning activities? In the next section, the detailed procedures regarding how large and massive data, where the individual course is a unity of analysis, were manipulated and transformed to extract common features of all courses considered as blended learning.

4. Methodology

4.1. Research context

The research context is a large private woman's university, located in Seoul, Korea. Around 22,200 students (16,100 undergraduate students, 6100 graduate students) are enrolled and around 1000 full-time faculty members are employed (the proportion of full-time faculty to part-time faculty is 35.8%). According to the annual report for self-evaluation of this university (2014), the institution takes special care of fulfilling the condition of more than 25% for small size class (less than 20 students per class) and less than 2.5% for large size class (more than 100 students per class). The small size classes (less than 20 students per class) were 901 (34.9%), and the large size classes (more than 100 students per class) were 98 (3.8%) out of 2579 classes in the spring semester of 2014. The rest of middle size courses were 1580 (61.3%).

In referring to a self-evaluation checklist based on the blended learning adoption framework devised by Graham et al. (2013), this institution is at a stage of adoption and early implementation (level 2). Although a uniform definition or policy on blended learning has not been widely shared among all university members, official documents have indicated that the university promotes blending (converging) online and offline learning environment to realize smart education. A sub-organization called the Institution of Teaching and Learning (ITL) is a designated sub-organization to support required infrastructure, contents development, and faculty/students interactions. As relatively recent action, iPad as a tool to support instructor's smart teaching was provided to 70% of full-time faculty members. Since 2011 faculty training programs to facilitate the class utilizing smart device and virtual classroom have been actively developed and implemented. As the most recent policy, the university decided to develop several K-MOOCs (Korean Massively Open Online Courses) which can be implemented to the facilitation of flipped learning as a format of bended learning for not only massive outside students but also university students.

This university has adopted Moodle, an open source learning platform and customized for the local service to this university since 2011. Like other LMSs, Moodle provides diverse functions such as learner profiles, learning materials, assignment submission, online tests/quiz, discussion boards, chatting tools, and file repositories. It also enables the management of multiple courses, discussion forums, and resources, and support diverse teaching and learning methods including project based instruction and collaborative learning. As Krüger, Merceron, and Wolf (2010) illustrated, the system contains three kinds of data elements, specified by tables to describe objects, interactions with learning objects, and association tables to describe links between objects. Diverse users such as lecturer, administrator, tutor, and student register the specific courses and the system support the management of courses with diverse functions such as quiz, discussion forum, wikis, and resources. LMS stores all the logs of users with time stamp and types of interactions such as view, modification, creation, attempt, and submit.

ITL operates this Moodle LMS. Professional staff in ITL support overall technical needs and provide pedagogical models so that instructors facilitate students' active online engagement. There has been increasing expectation and demand that faculty members utilize LMS more actively and smartly. For example, prior studies (Park & Jo, 2013, 2014) investigated what extent to which the students in this university use LMS and how they perceive the needs regarding the smart use of LMS. In the qualitative research, while participants perceived the usage of LMS as a useful tool to enrich their learning experience and improve the class management efficiency, they claimed that the level of LMS usage depends on the skills and efforts of instructors. With the adoption of analytics in 2013, ITL began reporting the statistics regarding the number of courses utilizing virtual classroom, the devices of users for using LMS (e.g., pad, phone, desktop), and users' weekly login patterns to promote the integration of LMS into the classes. The university annual evaluation report (2014) indicated the increased number of courses that are incorporating virtual classrooms.

Further, ITL has introduced faculty members who are implementing blended learning strategies creatively and enthusiastically as best practices to share with other members. In the past few years, to promote innovative pedagogical models and best practices, ITL began to use an automatic system to force all opened courses each semester to possess a virtual learning environment so that instructors, assistants, and students can use the prepared virtual space with built-in default functions such as announcement board or resource rooms. As a result, we found 4416 courses in LMS during the fall semester of 2013 and used them to conduct data-mining for this study.

4.2. Data mining process

According to Xu and Recker (2012), data-mining process entails pre-processing datasets, apply data mining algorithms to analyze the data, and post-process results. As the first step, data-preprocessing is especially important since it converts large, massive, and noisy dataset to a format to find meaningful patterns. The tasks for data-preprocessing include data cleaning, missing value imputation, data transformation, and data integration. In our case study, we conducted pre-processing task with the following three steps: data extraction, data cleaning, and elimination of missing values.

For the first *extraction* step, we collected all individual course (the unit of analysis) related data from two databases: 1) Course Management System (CMS), and 2) Learning Management System (LMS). Data from the CMS included course-related information indicating hierarchical categorizations of specific courses such as graduate vs. undergraduate, mandatory vs. selective, affiliated colleges and department. The LMS database included information on total members, log-in frequencies of each member, and log-data related to the teaching and learning activities of each member of the specific course. In the second *mapping* step, two datasets were combined by course ID as a key value. In addition, since there is a possibility to reveal the individual instructors' names and online behaviors left in the system due to the use of real course ID, the course ID was converted into numbers. The de-identification procedure was done to help minimize possible ethical issues. During this mapping process, 273 courses were canceled or closed. Therefore, we deleted these courses. Therefore, the final course number was 3838. Lastly, *elimination of missing values* was carefully considered because the dataset still had numerous zeros (0) and blanks in the log data. In the case of data set analyzed, most of the missing values regarding online activity were caused by little or no existence of the specific activity in the courses. For example, cases with total number of login by all members less than the total number of members and cases without lecture notes or resources were detected. Therefore, a total of 1472

Table 1
Descriptive statistics of 2639 courses.

Variables	Abbr.	Min	Median	Max	Mean	Mode	SD
Class size (# of members)	MEM	2	24	301	33	6	33.66
Login frequency per course	LF	8	710	21,414	1,399	75	1918
Average log-in frequency per person	ALF	4	30	71	40	12.5	33.01
Number of activity item	NAI	1	2	8	2.49	2	1.3

courses in these cases were further removed. As a result, a total of 2639 courses were left for our data analysis.

4.3. Data analysis and methods

Before entering into the initial data analysis step, we judged that the refined 2639 courses were still not in good shape to conduct a cluster analysis. In addition, such courses could not be even considered as blended courses in spite of their online activity traces in LMS log. Consequently, we used a two-phase systematic analysis approach. The *first* phase was focused on the investigation of the 2639 courses to explore online activity level and patterns. Descriptive statistics was mainly used for this phase study. Three variables [class size (number of students), login frequency, and activity diversity (number of activity items)] were reviewed to evaluate the overall quality of online activity. Ten activity-based variables (announcement, links, lecture notes, resources, Q&A, discussion forum, quiz, group project, Wikis, and assignment submission) were also reviewed.

In the *second* phase, by specifying the inclusion and exclusion criteria, 612 courses were selected to discover subgroups or clusters that were similar to each other on various online instructional interventions by using Latent Class Analysis (LCA) in Mplus Version 6. LCA is a statistical technique for multivariate categorical data to identify common patterns among a set of variables and classify individual unit of analysis into unobserved subgroups (Geiser, 2012). Several model fit indices were used to determine the optimal number of subgroups, including Akaike Information Criterion (AIC), Bayesian Information Criterion (BIC), the Lo–Mendell–Rubin Likelihood ratio test (Muthén, 2003), and entropy (Muthén & Muthén, 2000). The AIC and BIC are goodness-of-fit measures. Lower AIC and BIC values indicate better fitting models. Entropy is a standardized measure of the classification accuracy of class assignment. The higher the values are, the greater the precision of the classification (ranged from 0 to 1) is. The LMR-LRT is a test of statistical significance to compare a model with a given number of classes with a model with one fewer class. For example, if the LMR LRT for *c*-number model has significance at $p < .05$, the *c* class model is the best fit, while the *c-1* class model is rejected.

5. Results

5.1. Phase 1: overall patterns of online activity

As shown in Table 1, the class size of course varied, from 2 to 301 students in a class (Mean = 33, SD = 33.66). A basic log variable to grasp the overall level of online activity was *Login-Frequency* (*LF*). This variable was made by adding up all the login entries automatically registered whenever students of each course accessed a specific course's virtual classroom. However, as the number of students in a class gets bigger, the total login frequencies of specific course would be increased. Therefore, a new variable named *Average Log-In Frequency per person* (*ALF*) was made. For example, approximately 33 members logged in the virtual classroom an average of 1399 times per semester (15 weeks), resulting in frequency of 40 average logins per person. The results indicate very low online access. The high standard deviation values and L-Shape indicated heavy-tailed skewed distributions.

Another variable to estimate the overall patterns of online activity is the *Number of Activity Items* (*NAI*). While the login frequency variable can estimate the *activeness* of a virtual classroom, this variable can be used to measure the *diversity* of activity incorporated in the virtual classroom. In the present study, "*activity*" was specified in ten items based on the functions provided by the Moodle-based LMS. A framework showing the extractable log variables marked by circles (O) is presented in Table 2. A Moodle course is carried out by integrating the following two digital contents: *resources and activities* (Blin & Munro, 2008). While *resources* including web pages (html files) or text messages (txt files) were directly created (linked or posted) via the Moodle

Table 2
Data frame of the activity items and measurement scales.

Measurement scale / Activity items			# of uploads	# of replies or submits	# of short replies	# of views or downloads	Interaction type
1	Announcements	Instructor	0				
		Students					
2	Links	Instructor	0				"Instructor-led"
		Students					Mostly instructors uploaded posts, while students download and read.
3	Lecture notes	Instructor	0				
		Students					
4	Resources	Instructor	0			0	
		Students	0			0	"Interactive activities"
5	Q&A	Instructor	0	0	0	0	Anybody can post and share information.
		Students	0	0	0	0	
6	Discussion forum	Instructor	0			0	
		Students	0			0	
7	Quiz	Instructor	0				
		Students		0			
8	Group project	Instructor	0	0	0	0	"Instructor-guided, student-centered activities"
		Students	0	0	0	0	Given instructors' initiation, students work individually and/or in group.
9	Wiki	Instructor	0			0	
		Students	0			0	
10	Assignment submission	Instructor	0				
		Students		0			

Table 3
Descriptive statistics on online activity items (N = 2639).

	Activities	Courses that included each of these activities		Max	Mean	SD
		Frequencies	Percent (%)			
1	Announcements	1988	75.33	132	6.64	9.26
2	Links	126	4.77	72	0.32	2.57
3	Lecture notes	629	23.83	176	3.74	9.69
4	Resources	2038	77.23	596	11.87	21.49
5	Q&As	819	31.03	280	2.95	14.25
6	Discussion forum	77	2.92	2810	6.45	75.32
7	Quiz items	27	1.02	215	0.61	8.34
8	Group works	204	7.73	1612	17.52	88.42
9	Wikis	6	0.23	15	0.01	0.313
10	Assignment submissions	599	22.7	36	0.95	2.82

interface, other resources such as PowerPoint slides shows, word documents, Flash animations, videos and audio clips were externally developed and uploaded into the system. The activities allowed by Moodle are mostly *interactions* between instructor and students, among students, and between students and resources (contents). Consequently, all possible activities that could occur within this institution's virtual classrooms amounted to ten items listed in Table 2.

The first three items (announcement, links, and lecture notes) were mostly performed by the instructor and his or her teaching assistants. The fourth item (resources) indicated the sharing of any kind of contents uploaded not only by the instructor, but also by students themselves. In addition to resources, using Q&A boards or discussion forum required social interactions between students and instructors. Other activity items such as quiz, group project (using group board), collaborative writing via Wiki, and assignment submission required initial facilitation by instructors, but were sustained by students' active participation. To what extent the online activity occurred in a specific course could be measured by several different scales, such as the number of uploads, the number of replies or submits (in the case of assignments), the number of short replies (in the case of Q&A and Group Board), and the number of views or downloads (in the case of resources). The new variable NAI was created based on the framework in Table 3 and figuration of the ten activity items. Therefore, the possible maximum value of NAI was 10, while the minimum was 1. As shown in Table 1, the maximum value observed in the current dataset was 8, indicating that none of the classes utilized all ten activity items. Most courses incorporated two or three activity items into their virtual class.

The descriptive statistics of the ten variables are presented in Table 3. The dataset contained heavily skewed variables with too many "0 (zero)"s, implying that most courses did not incorporate diverse activity items. In other words, most blended learning courses were set up so that when the students are online they can view their course websites. However, those courses rarely provided active learning opportunities online. The maximum values of each activity were 132 announcements, 72 links, 176 lecture notes, 596 resources, 280 Q&A postings, 2810 discussion postings, and 1612 group board postings. Although this result itself is an evidence of significantly high presence of online, such extremely deviant values and considerably large amount of "0 (zero)" hindered further analysis such as classification or prediction due to the distorted estimation (Xu & Recker, 2012; Zimmerman, 1994). Therefore, two strategies were incorporated: case reduction and segmentation for the phase 2 study. Detailed explanations are introduced in the next section.

5.2. Phase 2: classification of blended courses

At the first step of phase 2, data segmentations and purposeful sampling were established to select the most representative ones of the blended course profiles. According to Patton (1990), "the logic and power of purposeful sampling lies in selecting information-rich cases for study in depth" (p. 169). Considering the skewed and heavy tailed distributions across online activity items, all items were converted into ordinal class values. For example, as shown in Table 4, the variables of the 2 most frequently used learning activities (e.g., resources and announcements) were converted with 4 levels by using the quantile score of each variable (25%, 50%, 75%, and 100%), while those of the least frequently used activities (e.g., Q&As, lecture notes, assignment submission, group work, links, discussion forums, quiz, and wikis) were redefined as binary classification (Yes—used, No—not used). This reveals that the most frequently used activities consist of generally required components of online instructional delivery, while the least frequently used activities are more likely associated with deeper learning for engaging student-centered learning experiences, as shown in Table 2. These results align with other research showing that most cases of blended learning only applied basic features of the LMS such as simply accessing course content, rather than advanced features

Table 4
Segmentation of variables.

Variables	Segmentation	Range	n	Min.	Max.	Mean	SD	Skewness	Kurtosis
Resources	1	8 ~ 11	124	8.00	401.00	23.00	25.60	8.27	98.17
	2	12 ~ 17	180						
	3	18 ~ 24	150						
	4	25 ~ 401	158						
Announcements	1	4 ~ 5	126	4.00	82.00	13.30	10.16	2.28	7.75
	2	6 ~ 9	148						
	3	10 ~ 15	164						
	4	16 ~ 82	174						
Q&A	1	0	277	0.00	280.00	7.81	25.73	6.99	57.92
	2	1 ~ 280	335						
Lecture notes	1	0	489	0.00	48.00	2.64	6.99	3.29	12.02
	2	1 ~ 48	123						
Assignment submission	1	0	408	0.00	24.00	1.28	3.09	3.91	18.13
	2	1 ~ 24	204						
Group works	1	0	507	0.00	1.00	0.17	0.38	1.75	1.05
	2	1~	105						
Links	1	0	581	0.00	1.00	0.05	0.22	4.11	14.93
	2	1~	31						
Discussion forums	1	0	588	0.00	1.00	0.04	0.19	4.76	20.72
	2	1~	24						
Quiz	1	0	604	0.00	1.00	0.01	0.11	8.60	72.11
	2	1~	8						
Wiki	1	0	608	0.00	1.00	0.01	0.08	12.28	149.23
	2	1~	4						

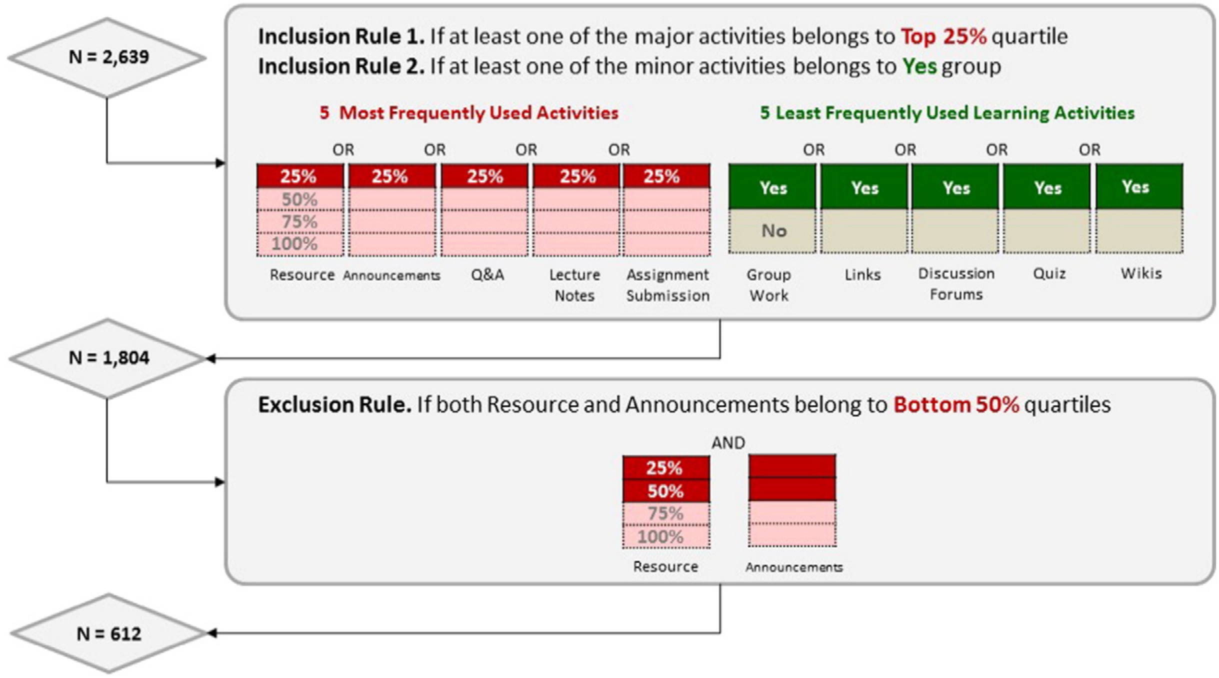
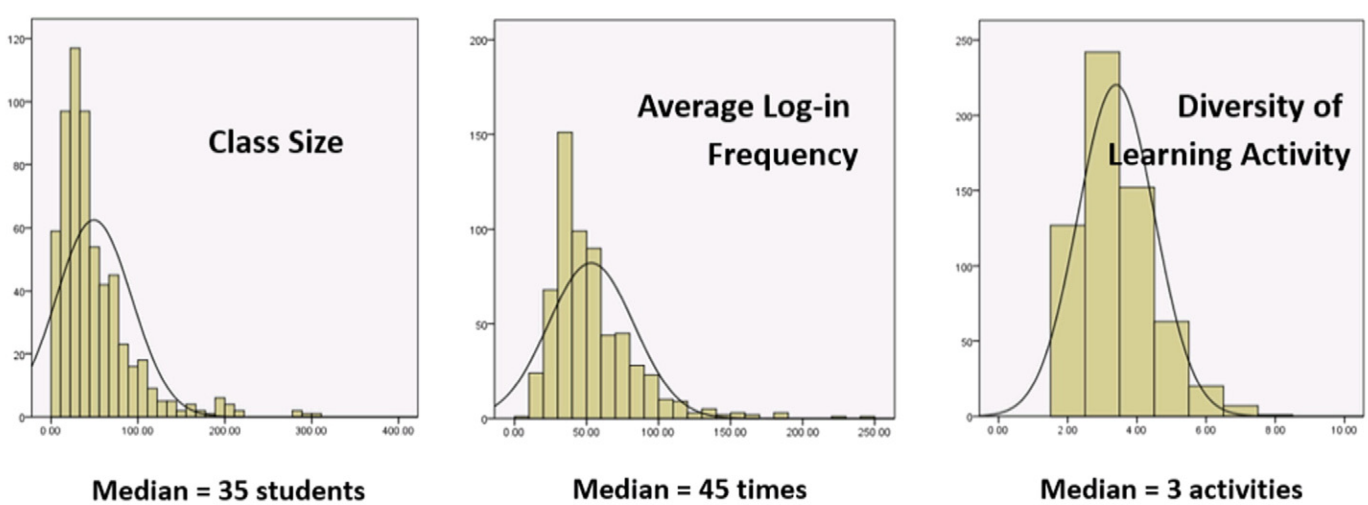


Fig. 2. Inclusion and exclusion criteria to reduce course from 2639 to 612.

such as using it in more engaging, customized, and collaborative ways (Dahlstrom et al., 2013).

After data segmentation, the inclusion and exclusion criteria were established as shown in Fig. 2. We only included courses that demonstrated top 25% usage of at least one major learning activity and those contained at least one of the minor learning activities. As a result, 1804 courses were selected to manually check the patterns of their online learning activities. We found that the majority of these courses still had significantly low usage that might confuse the data interpretation. Thus, 1192 courses belonging to the low 50% usage groups of the top two major learning activities such as resources and announcements were additionally excluded. Consequently, the sample of the phase two included 612 courses.

To review the general features of the selected 612 courses, descriptive statistics were conducted again. The majority of the courses were undergraduate-level ($n = 493$, 80%) courses designed as major-specific ($n = 245$, 40%) or liberal arts ($n = 181$, 30%) courses. As shown in Fig. 3, variables such as MEM, ALF, and NAI turned out to have normal distribution. By using the median score, we found that the distributions were almost equally divided into four quartiles. That



Course Type	Members	Case
Small	3~21	151
Medium	22~35	155
M-L	36~64	151
Large	65~301	155

Course Type	Log-in	Case
Rarely	8~32	137
Occasionally	33~44	166
Frequently	45~66	160
Very Frequently	67~245	149

Course Type	Activity	Case
Not diverse	2	137
Little diverse	3	166
Somewhat Diverse	4	160
Very diverse	5~8	149

Fig. 3. Histograms of MEM, ALF, and NAI ($n = 612$).

Table 5
Latent class model comparisons.

Number of classes	AIC	BIC	aBIC	Chi-square	LMR (p-value)	Entropy
2	6689	6817	6725	1172	121 ($p < .05$)	.671
3	6681	6875	6736	1148	37.53 ($p = .53$)	.610
4	6682	6943	6756	1042	27.97 ($p < .05$)	.666
5	6678	7005	6770	1013	41.99 ($p = 1.0$)	.661
6	6681	7075	6792	959	37.87($p = .77$)	.703

Note. AIC: Akaike Information Criterion (Lower = Better); BIC: Bayesian Information Criterion (Lower = Better); aBIC : the adjusted BIC (Lower = Better), LMR : Lo-Men-dell-Rueben (with good model fit indicated by $p < .05$), Entropy (Closer to 1 = Better).

is, the data sets of 612 blended courses present that around 35 students in a class logged in the virtual classroom at an average of 45 times and participated in three online activities.

To build a classification model as in the second step, a process to decide the number of latent class was carefully considered throughout the review of model fit. Table 5 and Fig. 4 provide the fit statistics for Model 1 (2-class model) through 5 (6-class model). The AIC values showed a decreasing pattern, whereas the BIC and adjusted BIC values showed an increasing pattern. Therefore, the LMR-LRT showed significant P-values with 2 and 4 classes, suggesting that these two models were the more appropriate solutions than 5 or 6 classes. Considering the fact that the 4-class model had less BIC values, it was concluded that the 4-class model had the best fit with the highest level of separation and the most distinctive patterns.

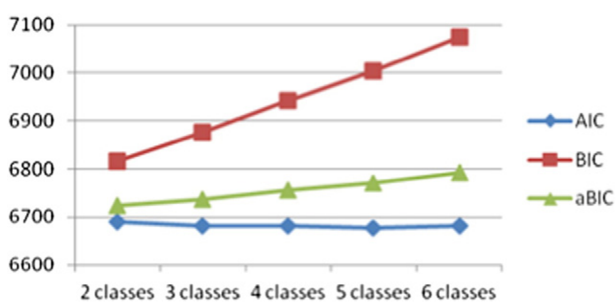


Fig. 4. Scree plot.

Within the 4-class model, we compared each class by using their profiles of online learning activities as shown in Fig. 5.

Cluster 1 was labeled *C* (*Communication or Collaboration*)*-focused BL Course* (n = 149, 24.3%), demonstrating significantly higher usage of Q&A (with .704 probability) and Group Works (with .291 probability). The majority of the courses were from the College of Social Sciences (n = 15, 10.1%), College of Education (n = 11, 7.4%), and Liberal Arts (n = 11, 7.4%) (See Table 6). A further analysis was performed to review the general features of online activity. Results are summarized in Table 7. Cluster 1 showed high level of Log-in frequency (45 or more access a semester = 74. 5%) as well as diverse online learning activities (4 or more activities = 76.8%). Cluster 2 was labeled *D* (*Delivery or Discussion*)*-focused BL Courses* (n = 110, 18.0%), showing higher usage of Lecture Notes (Probability = .757), Links (Probability = .256), and Discussion Forum (Probability = .130). This type of blended learning courses were from both College of Engineering (n = 11, 10.0%) and College of Social Sciences (n = 11, 10.0%). Class 2 also showed high levels of Log-in frequency (45 or more access a semester = 43.8%) and diversity in online learning activity (4 or more activities = 98.3%).

With the smallest portion, Cluster 3 was labeled *S* (*Sharing or Submission*)*-focused BL Course* (n = 44, 7.2%), indicating higher usage of Resources and Assignment Submission. This class showed relatively low level of Log-in Frequency (less than 44 access = 52.3%) and diversity in online learning activities (3 activities = 68.2%). This type of courses were mostly from the College of Education (n = 8, 18.2%) and College of Business Administration (n = 5, 11.4%). Overall, Cluster 4 showed lower levels of usage across online learning activities with the largest portion, thus was labeled I (*Inactive or Immature*)*-focused BL Course* (n = 309, 50.5%). Except resources and assignment submission, there was almost no usage of online learning activities. This type of courses was mostly from the College of Business Administration

6. Discussion and conclusion

In this study, as an approach of Academic Analytics, a classification of blended learning courses was attempted by mining the educational data sets derived from the Learning Management System (LMS) and institution's course database.

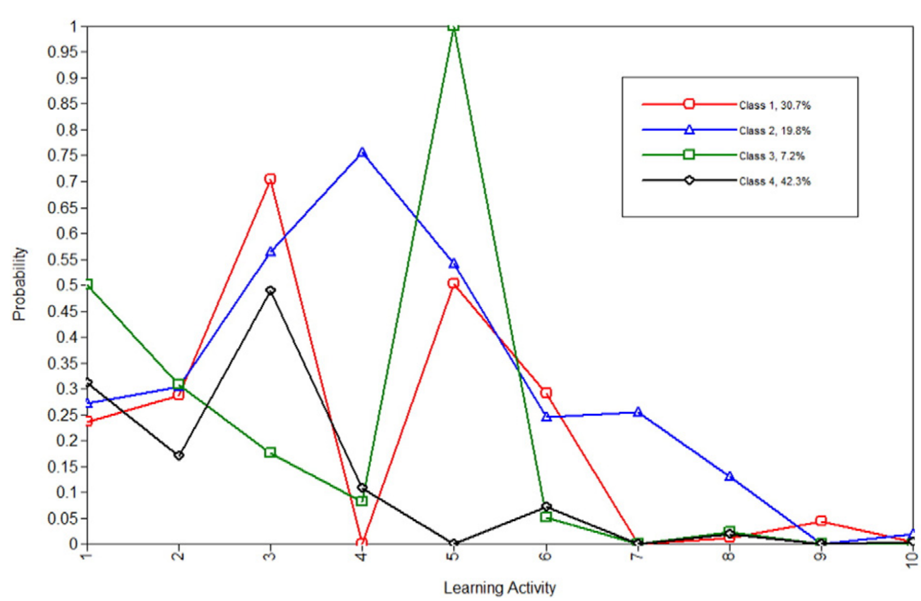


Fig. 5. Probability plots of the four classes based on the 10 online activity items. (1 = Resources, 2 = Announcements, 3 = Q&A, 4 = Lecture Notes, 5 = Assignment submission, 6 = Group Works, 7 = Links, 8 = Discussion Forums, 9 = Quiz, 10 = Wikis).

Table 6
Demographic characteristics of four clusters (N = 612).

Cluster	Level 1	Level 2	Level 3	n (%)
Cluster 1: 149 (24.3%)	Undergraduate: 131 (87.9%)	Liberal Arts: 57 (38.3%)	Mandatory	43 (28.9%)
			Selective	14 (9.4%)
		Major Specific: 66 (44.3%)	Social Sciences	15 (10.1%)
			Others.	51 (34%)
		Major Foundation: 8 (5.4%)		
	Graduate: 18 (12.1%)	General: 7 (4.7%)		
		Special: 11 (7.4%)	Education	6 (4.0%)
			Others	5 (3%)
				149 (100%)
Cluster 2: 110 (18.0%)	Undergraduate: 87 (79.1%)	Liberal Arts: 34 (30.9%)	Mandatory	28 (25.5%)
			Selective	6 (5.5%)
		Major Specific: 43 (39.1%)	Engineering	11 (10.0%)
			Social Sciences	11 (10.0%)
			Others	21 (19%)
		Major Foundation: 10 (9.1%)		
	Graduate: 23 (20.9%)	General: 15 (13.6%)		
		Special: 8 (7.3%)	Education	4 (3.6%)
			Others	4 (4%)
	110 (100%)			
Cluster 3: 44 (7.2%)	Undergraduate: 32 (72.7%)	Liberal Arts: 3 (6.8%)	Mandatory	2 (4.5%)
			Selective	1 (2.3%)
		Major Specific: 26 (59.1%)	Education	8 (18.2%)
			Business Administration	5 (11.4%)
			Others	1 (2.3%)
		Major Foundation: 3 (6.8%)		
	Graduate: 12 (27.3%)	General: 5 (11.4%)		
		Special: 7 (15.9%)	Education	3 (6.8%)
			Clinical Health Sciences	3 (6.8%)
			Others	13 (30%)
				44 (100%)
Cluster 4: 309 (50.5%)	Undergraduate: 243 (78.6%)	Liberal Arts: 87 (28.2%)	Mandatory	68 (22.0%)
			Selective	19 (6.1%)
		Major Specific: 110 (35.6%)	Business Administration	23 (7.4%)
			Social Sciences	20 (6.5%)
			Liberal Arts	17 (5.5%)
			Education	16 (5.2%)
			Natural Sciences	12 (3.9%)
			Others	22 (7%)
		Major Foundation: 46 (14.9%)		
	Graduate: 66 (21.4%)	General: 32 (10.4%)		
		Special: 34 (11.0%)	Education	13 (4.2%)
			Others	21 (7%)
				309 (100%)

Table 7
General features of 4 Clusters of BL courses (N = 612).

Variables	Cluster 1 Type C (n = 149, 24.3%)	Cluster 2 Type D (n = 110, 18.0%)	Cluster 3 Type S (n = 44, 7.2%)	Cluster 4 Type I (n = 309, 50.5%)	F
Class size					
3–21	27 (18.1%)	25 (22.7%)	**17 (38.6%)**	82 (26.5%)	12.314
22–35	38 (25.5%)	30 (27.3%)	12 (27.3%)	75 (24.3%)	
36–64	**43 (28.9%)**	25 (22.7%)	10 (22.7%)	73 (23.6%)	
65–301	41 (27.5%)	30 (27.3%)	5 (11.4%)	79 (25.6%)	
Log-in frequency					
8–32	18 (12.1%)	7 (6.4%)	9 (20.5%)	103 (33.3%)	**174.176****
33–44	20 (13.4%)	8 (7.3%)	**14 (31.8%)**	**124 (40.1%)**	
45–66	**58 (38.9%)**	38 (34.5%)	13 (29.5%)	51 (16.5%)	
67–245	53 (35.6%)	**57 (9.3%)**	8 (18.2%)	31 (10.0%)	
Number of activity items					
2	0 (0.0%)	0 (0.0%)	0 (0.0%)	127 (41.1%)	**452.479****
3	36 (24.2%)	14 (12.7%)	**30 (68.2%)**	**162 (52.4%)**	
4	**86 (57.7%)**	36 (32.7%)	13 (29.5%)	17 (5.5%)	
5–8	27 (18.1%)	**60 (54.5%)**	1 (2.3%)	3 (1.0%)	

Note: $^{**}p < .01$.

Our results revealed that online behavior data tracked from LMS indicated the adoption level and patterns of blended learning implementation effectively. The dataset from one semester presented disappointingly low engagement of online activity compared to what the institution anticipated. Throughout multiple times of data mining process, 2639 out of 4416 courses (60%) were entered to phase 1 study to see to what extent online activity was incorporated for academic courses in practice. The dataset of 2639 courses presented a pattern of low usage of LMS from the most courses and very active incorporation of online activity from only a few classes. Such result indicated “adoption and early implementation” (level 2) in the blended learning adoption spectrum (Graham et al., 2013). Although this institution has indicated the willingness of promoting blended learning strategies throughout the university official document, more structured policies and supports need to be prepared for institutional-wide BL implementation and culture. For example, as mentioned in previous studies (Graham et al., 2013; Wallace & Young, 2010), the policies indicating clear reduction of f2f classroom hours by online teaching components should be provided and guided to faculty members as a strategy to promote blended learning. After this case study was conducted, we had a workshop consisting of staffs in ITL and research team in order to disclose the results of this study and discuss the future direction for BL strategy in this university. While the actual policy was not implemented yet, university staff members agreed on the necessity to adopt more powerful policy indicating the reduction in scheduled classroom time by 25 to 50% which was considered as key feature of BL success in a previous case study (Garrison & Vaughan, 2013). If such policies are implemented and university database records the proportion of online teaching hours in all courses, more sophisticated clustering of blended learning can be also possible. Since this study only observed the behavioral part of online learning activity, precise identification of the representative type of blended learning was a challenging work. Nevertheless, we believe the data-mining algorithm utilized for this study contributes to highlighting courses that present enthusiastic incorporation of online activity. We recommend an in-depth analysis that examines planned instructional methods of such courses in both online and offline side.

This study clustered 612 courses that were purposefully sampled with cases presenting relatively active online activity. Latent Class Analysis identified four types of blended course with C-D-S-I (C: Communication or Collaboration, D: Delivery or Discussion, S: Sharing or Submission, and I: Inactive or Immature). While 50% of courses fell into Type I, the rest of the courses were divided into Type C (24.3%), Type D (18%), or Type S (7.2%). The results revealed that students in Type C courses participated more in small-group project online as well as conversations with instructor and/or peers in terms of course requirements. Since students can have greater time flexibility in blended learning courses, they can reflect and refer what they learned from face-to-face instruction to collaborate through online group work to engage in authentic problem solving. This also allows instructors to address the different needs of individual groups and help them to truly engage in the subject based on their interests and depth of understanding, remaining “student-centered”. Students in Type D courses had online discussion as a substantial learning activity of their courses, which was designed to support students share experiences, insights, and perspectives related to course topics without time and place constraints. Through online discussion, students who are not active in class can be more "vocal" and this eventually helps create a stronger class community both in class and online. Unlikely Types C and D, it revealed that Type S provided somewhat less student-centered activities online, such as resource links and assignment submission menu. This type does not alter the traditional instruction, but just uses online components to supplement classroom lectures and learning materials. To increase the effectiveness of course logistics, instructors in Type S may carefully plan to compliment classroom activities, but not reduce face-to-face meeting times. Therefore, the order of C-D-S-I reflects the spectrum from active to passive engagement of online technology and student-centered paradigm, which can be also aligned with other adoption-based frameworks of blended learning (e.g., Graham et al., 2013; Francis & Raftery, 2005).

Further, we carefully forecast that the classes in the future university will be clustered in more divergent ways beyond the linear and convergent framework suggesting the direction from immature to mature status. Consequently, it is necessary to develop a model that fits to the ecological nature of diverse classes, ranging from small-size problem-based BL classes to large-size flipped BL classes, and the advanced technologies that support such diverse pedagogies. In this vein, our data-driven approach differed from the theoretically or conceptually classified types (Francis & Raftery, 2005; Graham et al., 2013; Margulieux et al., n.d; Singh, 2003). The C-D-S-I classification model reflects the phenomenon of current blended learning status inductively in this institution. Also, it is useful to develop more advanced predictive model and sophisticated intervention provision. For example, in prediction model in cases of type D, variables to measure the number of downloads or visits on discussion board might be an important factor for their academic success, while the courses of type C might need to consider variables such as the number of questions to make or the time to get responses from the instructor. In a similar vein, courses of type S would require considering variables such as the number of uploading posts or the number of assignment submission rather than downloading or visiting on board. In other words, different types of courses require different important variables to be included in the prediction and feedback-recommendation model. It was an only and critical limitation that this study did not include students' learning achievement variable or instructors' course evaluation results for the cluster analysis due to research ethical issues.

The results of this study suggest us to invest developing LMS in a strategic way. We found that most courses did not incorporate diverse activity items. The most frequently utilized activity item was resources, announcement, Q&A's, lecture notes, and assignment submission, while activities such as group works, quiz, Wikis, and discussion forums were the least utilized. Also, this study revealed dominantly affiliated college and department in four subtypes. Although such information can be useful as it is, it should be also remarked that distinctive patterns were not found to conduct cross-disciplinary planning of BL. Therefore, this study recommends a further investigation of representative pedagogical models and best practices depending on academic disciplines and customizes the LMS features based on the subtypes of blended courses. A university-wide course transformation initiative through blended learning (Garrison & Vaughan, 2013) is one of the critical dimensions for instructional endeavors for student success. Further examinations based on the findings of the present study might provide guidance in how to diagnose student performance online and how to develop strategic objectives for blended learning. However, this study depended on a dataset of a single university that has unique sociocultural characteristics. Since most courses in this Korean institution adopts a relative evaluation policy, which implies that top few students get 'A' mark. Consequently, students participate in activities that only relate to evaluation and tend to compete with each other to achieve a good final mark. In spite of the instructional efforts made by few enthusiastic instructors, such evaluation policy and cultural circumstances may have influence on students' LMS usage patterns and activity participation.

This study was meaningful as an institution-level investigation that enables us to present real phenomenon (even though the results were somehow disappointing) that might be hard to detect in individual course-level experimental research context depending on typically self-reported or interruptive data and many exogenous variables. We believe that the data-driven approach adopted in this study could be a seminal contribution to university staffs that need to monitor the whole status and to make better decisions based on the understanding of various contextual variables. This study can be further developed by re-examining the suggested four clusters at different higher education context. It is needed to enhance ecological validity and diagnostic sensitivity through the latent growth modeling over time.

References

Arnold, K. E. (2010). Signals: Applying academic analytics. *Educause Quarterly*, *33*(1), n1 (Available from http://er.educause.edu/articles/2010/3/signals-applying-academic-analytics).

Baepler, P., & Murdoch, C. J. (2010). Academic analytics and data mining in higher education. *International Journal for the Scholarship of Teaching and Learning*, *4*(2), 17.

Beisser, S. R., & Steinbronn, P. (2002). Hybrid online coursework to enhance technology competencies of school principals. *Computers in the Schools*, *19*(3–4), 191–205.

Bishop, J. L., & Verleger, M. A. (2013). The flipped classroom: A survey of the research. *Paper presented at the ASEE national conference proceedings*. Atlanta, GA: American Society for Engineering Education.

Blin, F., & Munro, M. (2008). Why hasn't technology disrupted academics' teaching practices? Understanding resistance to change through the lens of activity theory. *Computers & Education*, *50*(2), 475–490.

Cameron, B. (2003). The effectiveness of simulation in a hybrid and online networking course. *TechTrends*, *47*(5), 18–21.

Campbell, J. P., DeBlois, P. B., & Oblinger, D. G. (2007). Academic analytics: A new tool for a new era. *Educause Review*, *42*(4), 40.

Carroll, B., & Hsu, L. (2003). *Going hybrid: online course components increase flexibility of on-campus courses (Online Classroom)* , 4–7.

Carvalho, A., Areal, N., & Silva, J. (2011). Students' perceptions of Blackboard and Moodle in a Portuguese university. *British Journal of Educational Technology*, *42*(5), 824–841.

Cooner, T. S. (2010). Creating opportunities for students in large cohorts to reflect in and on practice: Lessons learnt from a formative evaluation of students' experiences of a technology-enhanced blended learning design. *British Journal of Educational Technology*, *41*(2), 271–286.

Dahlstrom, E., Brooks, D. C., & Bichsel, J. (2014). *The current ecosystem of learning management systems in higher education: Student, faculty, and IT perspectives: Research report.* Louisville, CO: ECAR (September 2014. Available from http://www.educause. edu/ecar. 2014 Educause. CC by-nc-nd).

Dahlstrom, E., Walker, J., & Dziuban, C. (2013). *ECAR study of undergraduate students and information technology, 2013.* Boulder, CO: Educause Center for Applied Research.

Donnelly, R. (2010). Harmonizing technology with interaction in blended problem-based learning. *Computers & Education*, *54*(2), 350–359.

Dziuban, C. D., Hartman, J. L., & Moskal, P. D. (2004). Blended learning. *EDUCAUSE Center for Applied Research Bulletin*, *7*(1), 12.

Francis, R., & Raftery, J. (2005). Blended learning landscapes. *Brookes eJournal of learning and teaching*, *1*(3), 1–5.

Garrison, D. R., & Kanuka, H. (2004). Blended learning: Uncovering its transformative potential in higher education. *The Internet and Higher Education*, *7*(2), 95–105.

Garrison, D. R., & Vaughan, N. D. (2013). Institutional change and leadership associated with blended learning innovation: Two case studies. *The Internet and Higher Education*, *18*, 24–28.

Geiser, C. (2012). *Data analysis with Mplus.* Guilford publications.

Goldstein, P. J., & Katz, R. N. (2005). *Academic analytics: The uses of management information and technology in higher education.* Educause.

Graham, C. R. (2006). Blended learning systems. *CJ Bonk & CR Graham, The handbook of blended learning: Global perspectives, local designs* (Pfeiffer).

Graham, C. R., Woodfield, W., & Harrison, J. B. (2013). A framework for institutional adoption and implementation of blended learning in higher education. *The Internet and Higher Education*, *18*, 4–14.

Halverson, L. R., Graham, C. R., Spring, K. J., Drysdale, J. S., & Henrie, C. R. (2014). A thematic analysis of the most highly cited scholarship in the first decade of blended learning research. *The Internet and Higher Education*, *20*, 20–34.

Hofmann, J. (2006). Why blended learning hasn't (yet) fulfilled its promises: Answers to those questions that keep you up at night. *The handbook of blended learning: Global perspectives, local designs* (pp. 27–40).

Krüger, A., Merceron, A., & Wolf, B. (2010). *A data model to ease analysis and mining of educational data.* Paper presented at the EDM.

Margulieux, L. E., Bujak, K. R., McCracken, W. M., & Majerich, D. (n.d). Hybrid, blended, flipped, and inverted: Defining terms in a two dimensional taxonomy.

Mostow, J., Beck, J., Cen, H., Cuneo, A., Gouvea, E., & Heiner, C. (2005). An educational data mining tool to browse tutor-student interactions: Time will tell. *Paper presented at the* *proceedings of the workshop on educational data mining, National Conference on Artificial Intelligence*.

Muthén, B. (2003). *Statistical and substantive checking in growth mixture modeling: Comment on Bauer and Curran (2003).*

Muthén, B., & Muthén, L. K. (2000). Integrating person-centered and variable-centered analyses: Growth mixture modeling with latent trajectory classes. *Alcoholism: Clinical and Experimental Research*, *24*(6), 882–891.

Nichols, M. (2008). Institutional perspectives: The challenges of e-learning diffusion. *British Journal of Educational Technology*, *39*(4), 598–609.

Norberg, A., Dziuban, C. D., & Moskal, P. D. (2011). A time-based blended learning model. *On the Horizon*, *19*(3), 207–216.

Park, Y., & Jo, I. (2014). Need analysis for learning analytics dashboard in LMS: Applying activity theory as an analytic and design tool. *Journal of Educational Technology*, *30*(2), 221–258.

Park, Y., & Jo, I. -h. (2013). Smart use of LMS in higher education: Viewing students' perceptions in a framework of activity theory. *Paper presented at the International Conference of Educational Technology, Sejong University*.

Patton, M. Q. (1990). *Qualitative evaluation and research methods.* SAGE Publications, inc.

Poon, J. (2013). Blended learning: An institutional approach for enhancing students' learning experiences. *Journal of online learning and teaching*, *9*(2), 271–288.

Porter, W. W., Graham, C. R., Spring, K. A., & Welch, K. R. (2014). Blended learning in higher education: Institutional adoption and implementation. *Computers & Education*, *75*, 185–195.

Rogers, E. M. (2010). *Diffusion of innovations.* Simon and Schuster.

Ross, B., & Gage, K. (2006). Global perspectives on blended learning: Insight from WebCT and our customers in higher education. *The handbook of blended learning: Global perspectives, local designs* (pp. 155–168).

Shank, P. (2011). *The online learning idea book: Proven ways to enhance technology-based and blended learning. Vol. 2*, John Wiley & Sons.

Siemens, G., & Long, P. (2011). Penetrating the fog: Analytics in learning and education. *Educause Review*, *46*(5), 30–32.

Singh, H. (2003). Building effective blended learning programs. *Educational Technology*, *43*(6), 51–54.

Tuckman, B. W. (2002). Evaluating ADAPT: A hybrid instructional model combining web-based and classroom components. *Computers & Education*, *39*(3), 261–269.

Vaughan, N. D. (2010). A blended community of inquiry approach: Linking student engagement and course redesign. *The Internet and Higher Education*, *13*(1), 60–65.

Wallace, L., & Young, J. (2010). Implementing blended learning: Policy implications for universities. *Online Journal of Distance Learning Administration*, *13*(4).

Xu, B., & Recker, M. (2012). Teaching analytics: A clustering and triangulation study of digital library user data. *Journal of Educational Technology & Society*, *15*(3).

Zimmerman, D. W. (1994). A note on the influence of outliers on parametric and non-parametric tests. *The Journal of General Psychology*, *121*(4), 391–401.