Looking into the black box of university classrooms through the windows of Moodle LMS

LAPA
LEARNING ANALYTICS FOR PREDICTION & ACTION

Yeonjeong Park, Ph.D.
(Honam University)

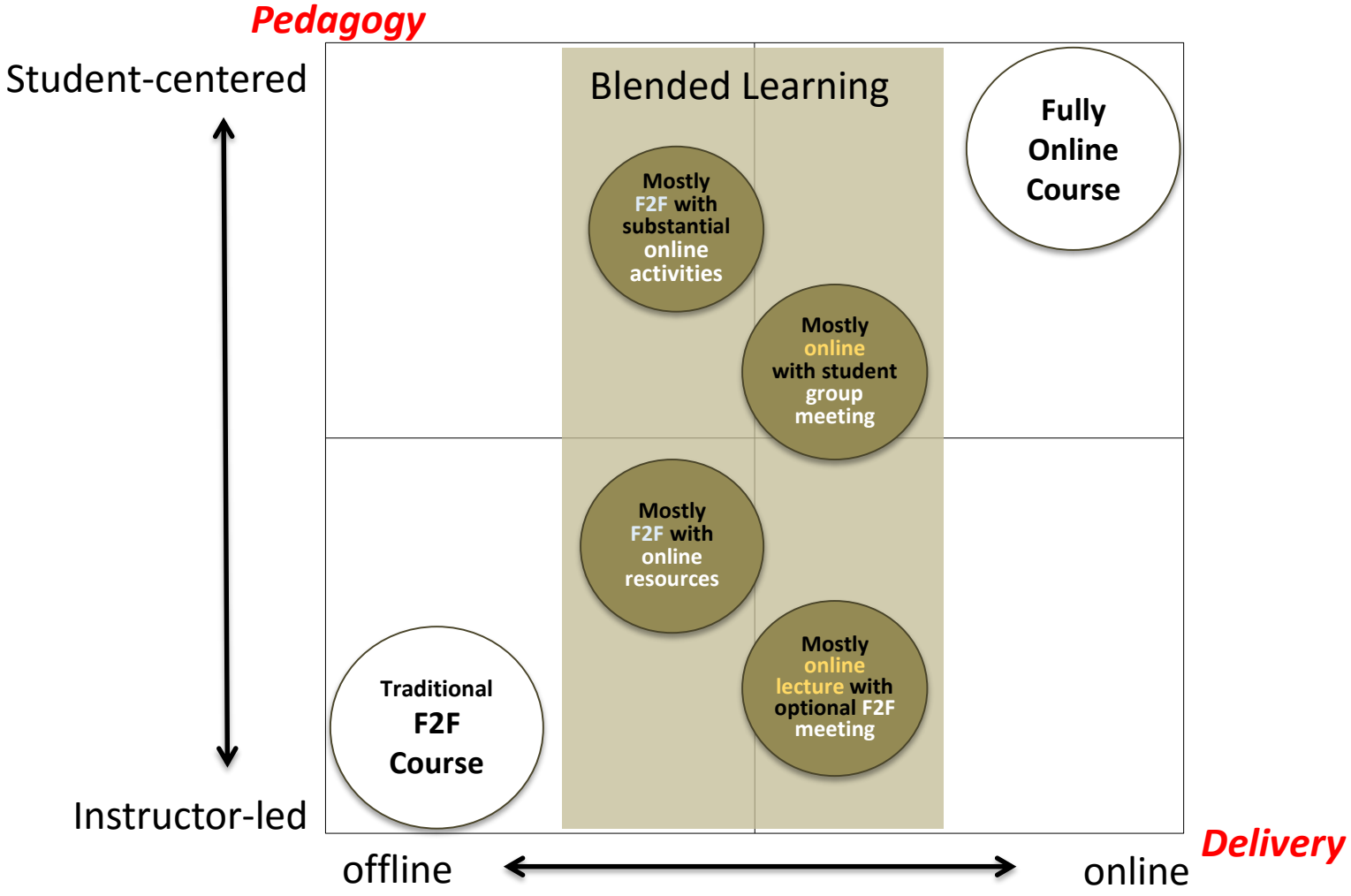
Contents

01. Backgrounds

LAPA
LEARNING ANALYTICS FOR PREDICTION & ACTION

Backgrounds

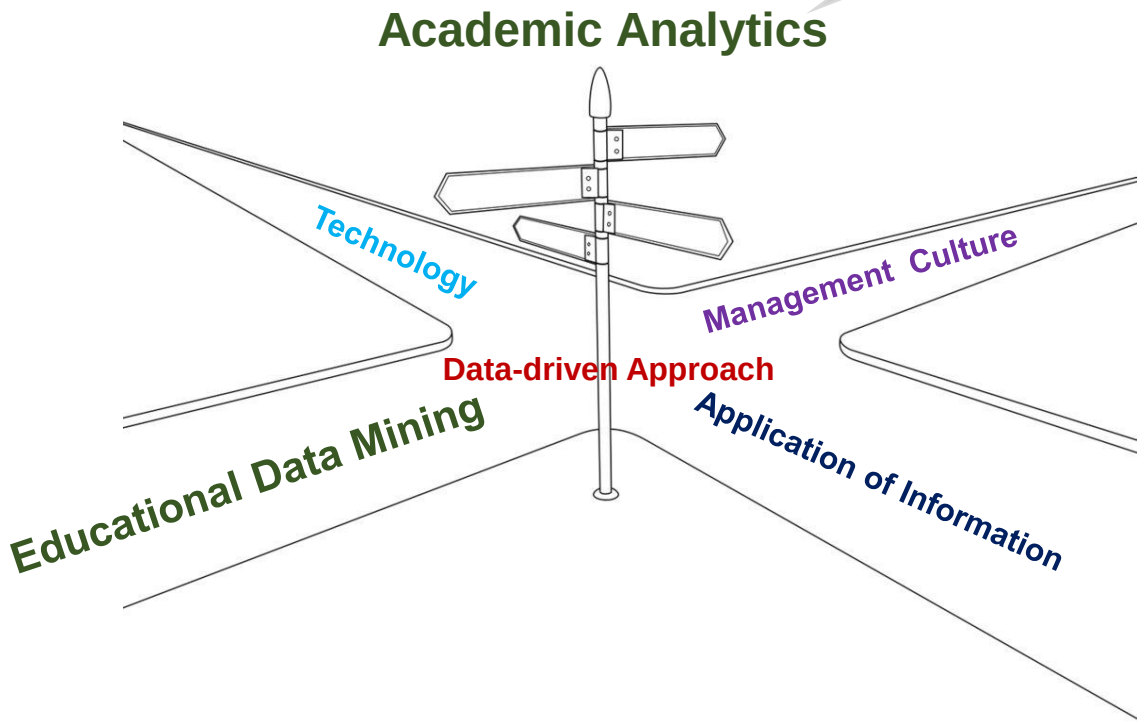
1) Blended Learning in Higher Education



Backgrounds

2) Academic Analytics

Guides course redesign and to implement evidence-based decision in higher education



- Derived from **business intelligence** (Goldstein & Katz, 2005)
- 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 (Campbell, DeBlois, and Oblinger, 2007)
- Examples: Early Warning System such as Course Signal of Purdue Univ.

Backgrounds

2) Academic Analytics

A speculative prospecting for riches **to find hidden patterns without a preconceived hypothesis**

1

Educational Data Mining

A scientific, hypothesis-driven approach using a particular dataset **to solve a practical academic problem**

Academic Analytics

Reflects the role of data analysis at **an institutional level**

2

Learning Analytics

Centers on the **learning process** including analyzing relationship between learner, content, institution, and education

1. 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.
2. Siemens, G., & Long, P. (2011). Penetrating the fog: Analytics in learning and education. Educause Review, 46(5), 30–32.

02. Research Purposes

LAPA
LEARNING ANALYTICS FOR PREDICTION & ACTION

Research Purposes and questions

Level and Patterns of Online Activity left in Moodle LMS

1) To what extent are university instructors and students using the LMS for their teaching and learning?

2) What are the most and least used activities in the LMS across diverse courses and colleges?

Classification of Blended Courses

1) What usage patterns and clusters emerge across blended learning courses, when mining students' participation data related to online learning activities?

2) What are the demographic and instructional characteristics of blended learning courses?

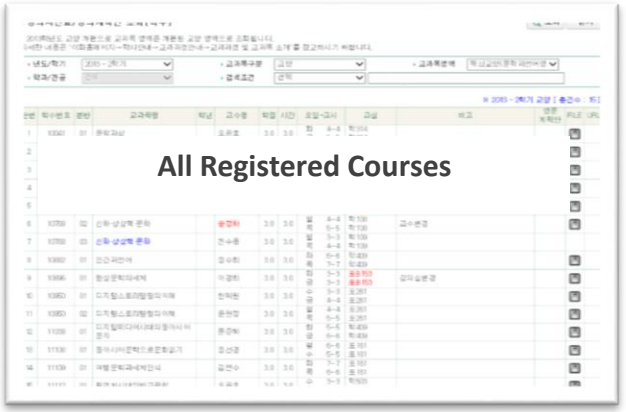
03. Methods

LAPA
LEARNING ANALYTICS FOR PREDICTION & ACTION

03 Methods

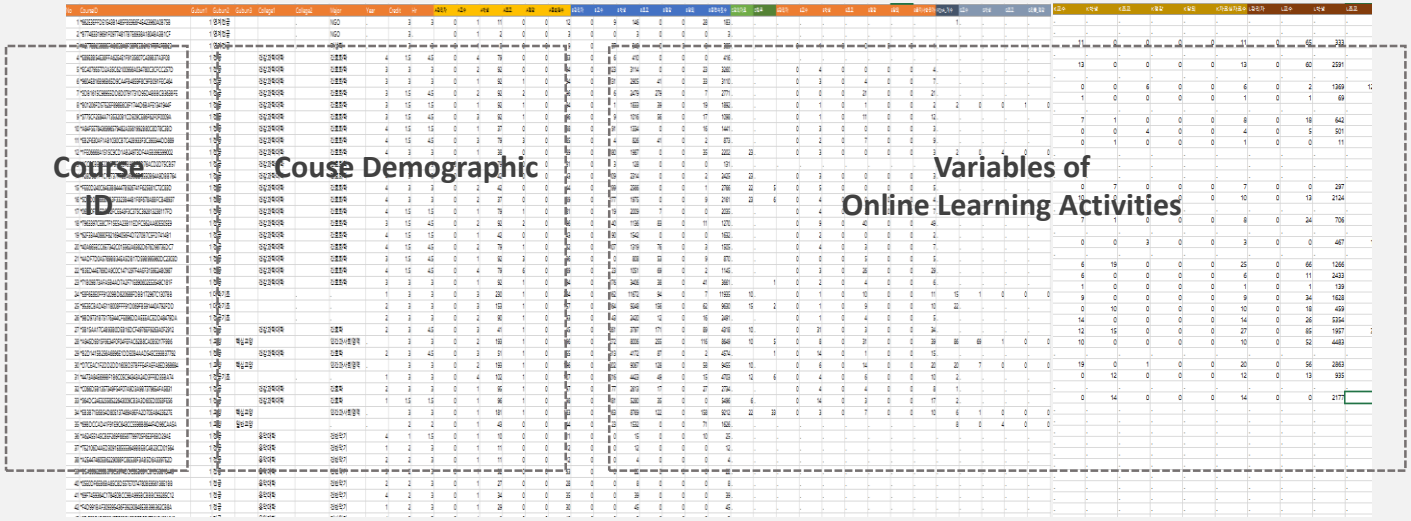
1) Research Contexts

Academic Affairs Office



Moodle-based LMS

One Large Data Set



Methods

1) Research Contexts

Measurement Scales of 10 Learning Activities

Activity Items (Variable) \ Measurement Scale			Number of uploads	Number of replies or submits	Number of short replies	Number of views or downloads	Interaction Type
1	**Announcements**	Instructor	O				**Instructor-led activity:** Mostly instructors upload posts, while students download and read.
		Students					
2	**Links**	Instructor	O				
		Students					
3	**Lecture Notes**	Instructor	O				
		Students					
4	**Resources**	Instructor	O			O	**Interactive activities:** Anybody can post and share information
		Students	O			O	
5	**Question & Answer**	Instructor	O	O	O	O	
		Students	O	O	O	O	
6	**Discussion Forum**	Instructor	O			O	**Instructor-guided, Student-centered Activities**: Given instructors' initiation students work individually and/or in group.
		Students	O			O	
7	**Quiz**	Instructor	O				
		Students		O			
8	**Group Project**	Instructor	O	O	O	O	
		Students	O	O	O	O	
9	**Wikis**	Instructor	O			O	
		Students	O			O	
10	**Assignment Submission**	Instructor	O				
		Students		O			

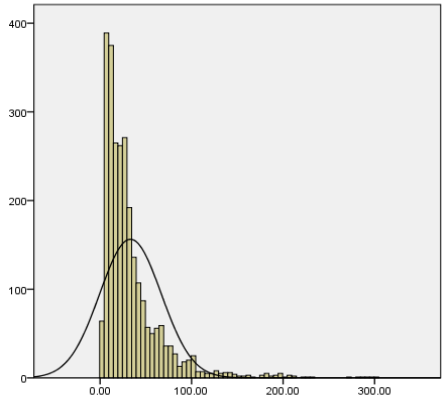
Methods

2) Descriptive Statistics (n=2,639)

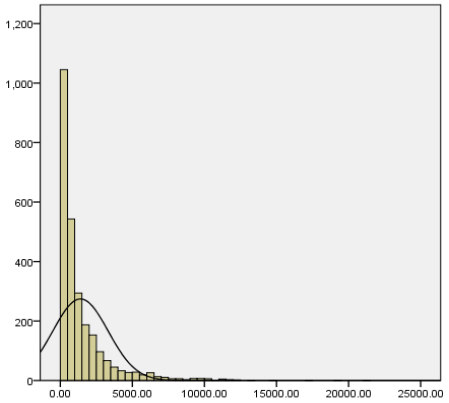
Descriptive Statistics of 2, 639 Blended Learning Courses

Variables	Min	Max	Mean	SD
Class Size (# of students)	2	301	33.00	33.66
Log-in Frequency per Course	8	21,414	1,398.62	1,918.00
Average Log-in Frequency per Student	2	375	39.75	33.01

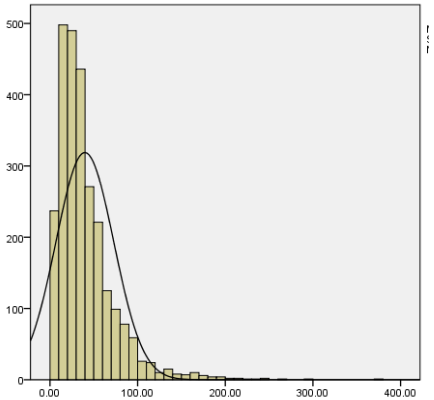
Very Skewed Heavy-tailed Distributions



Class Size



Log-in Frequency Per Course

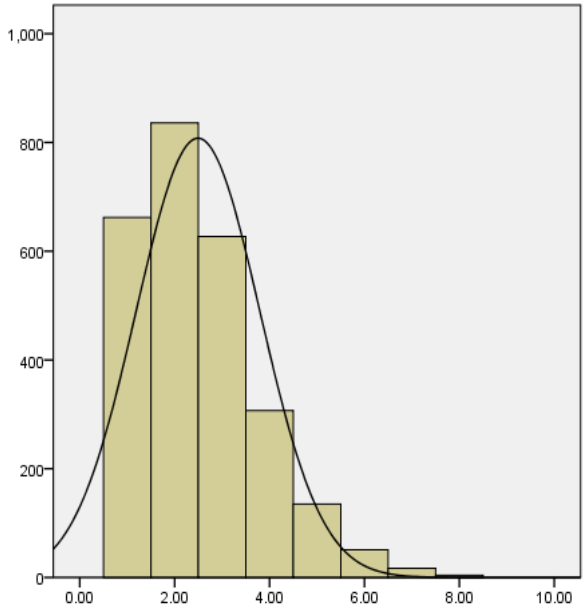


Log-in Frequency Per Student

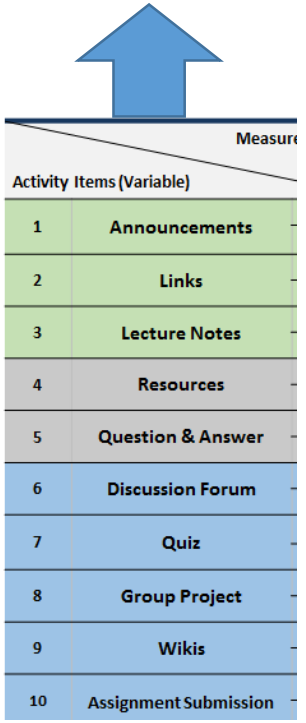
Methods

2) Descriptive Statistics (n=2,639)

Variable	Min	Max	Mean	SD
Number of Activity Items	1	8	**2.49**	1.3



Number of Activity Items



Methods

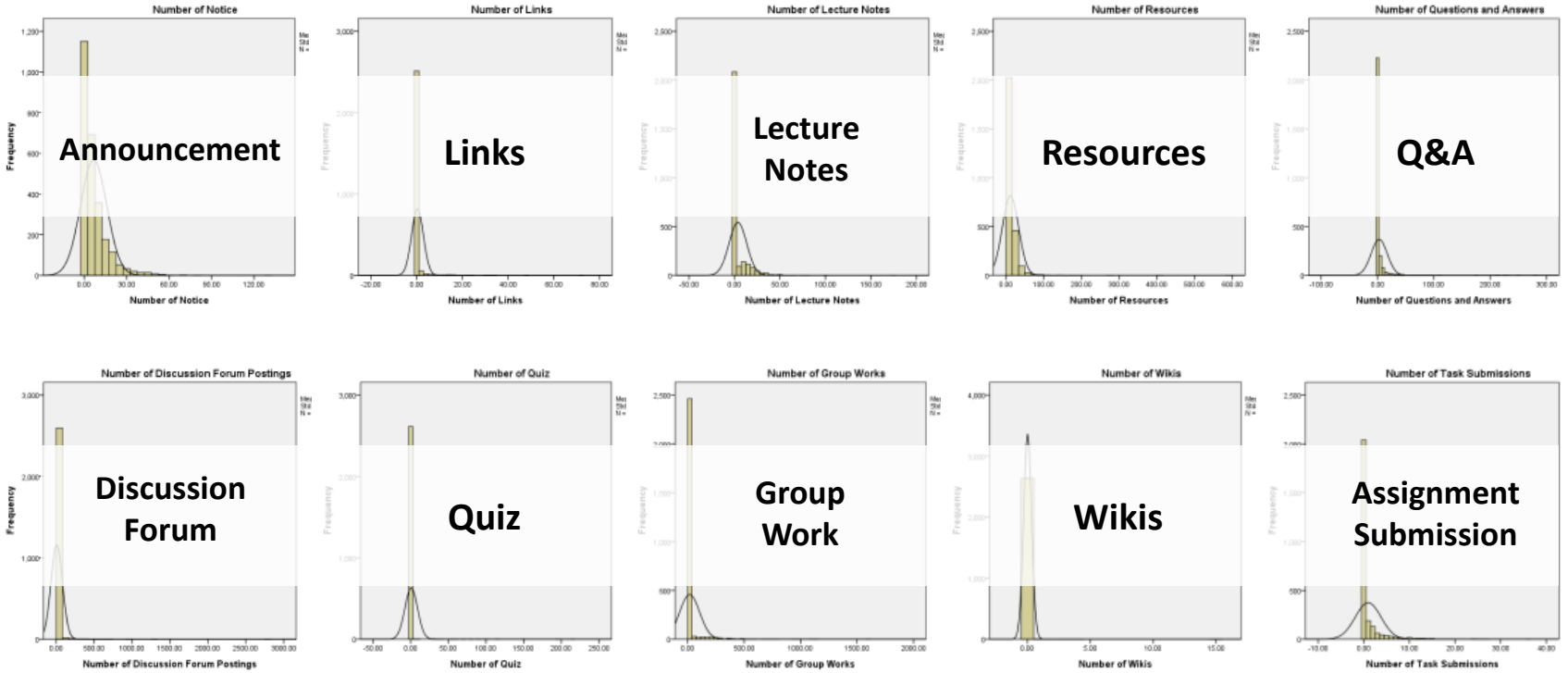
2) Descriptive Statistics (n=2,639)

	Learning Activity	Min	Max	Mean	SD	Skewness	Kurtosis
1	Number of Announcements	0	132	6.64	9.26	3.208	20.154
2	Number of Links	0	72	.32	2.57	14.971	312.540
3	Number of Lecture Notes	0	176	3.74	9.69	5.158	51.865
4	Number of Resources	0	596	11.87	21.49	12.222	263.560
5	Number of Q&A	0	280	2.95	14.25	12.089	183.950
6	Number of Discussion Postings	0	2810	6.45	75.32	24.442	788.773
7	Number of Quiz Items	0	215	.61	8.34	17.818	366.003
8	Number of Group Works	0	1612	17.52	88.42	8.226	91.714
9	Number of Wikis	0	15	.01	.313	42.921	2005.642
10	Number of Assignment Submissions	0	36	.95	2.82	4.972	32.731

Methods

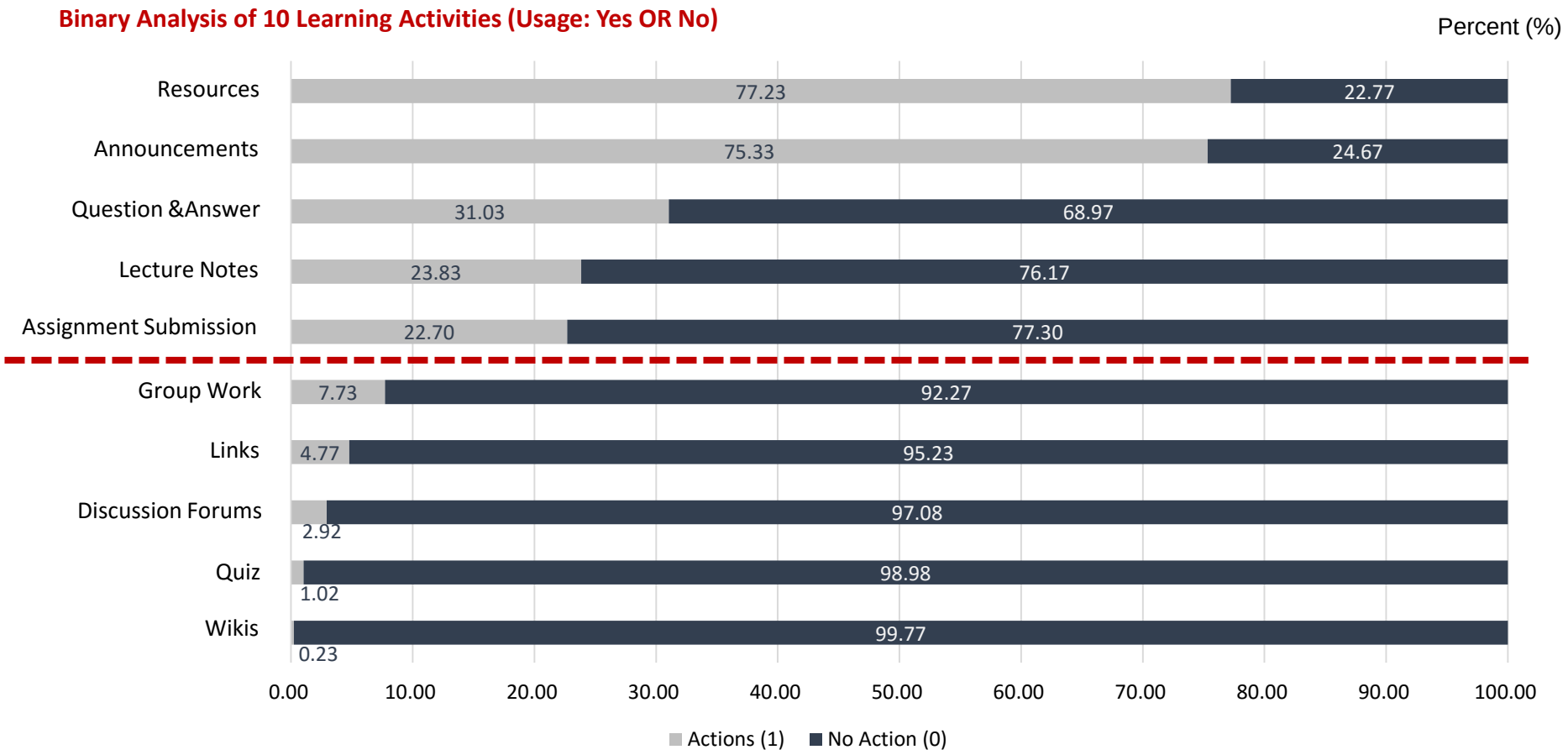
2) Descriptive Statistics (n=2,639)

Overall, very Skewed Heavy-tailed Distributions



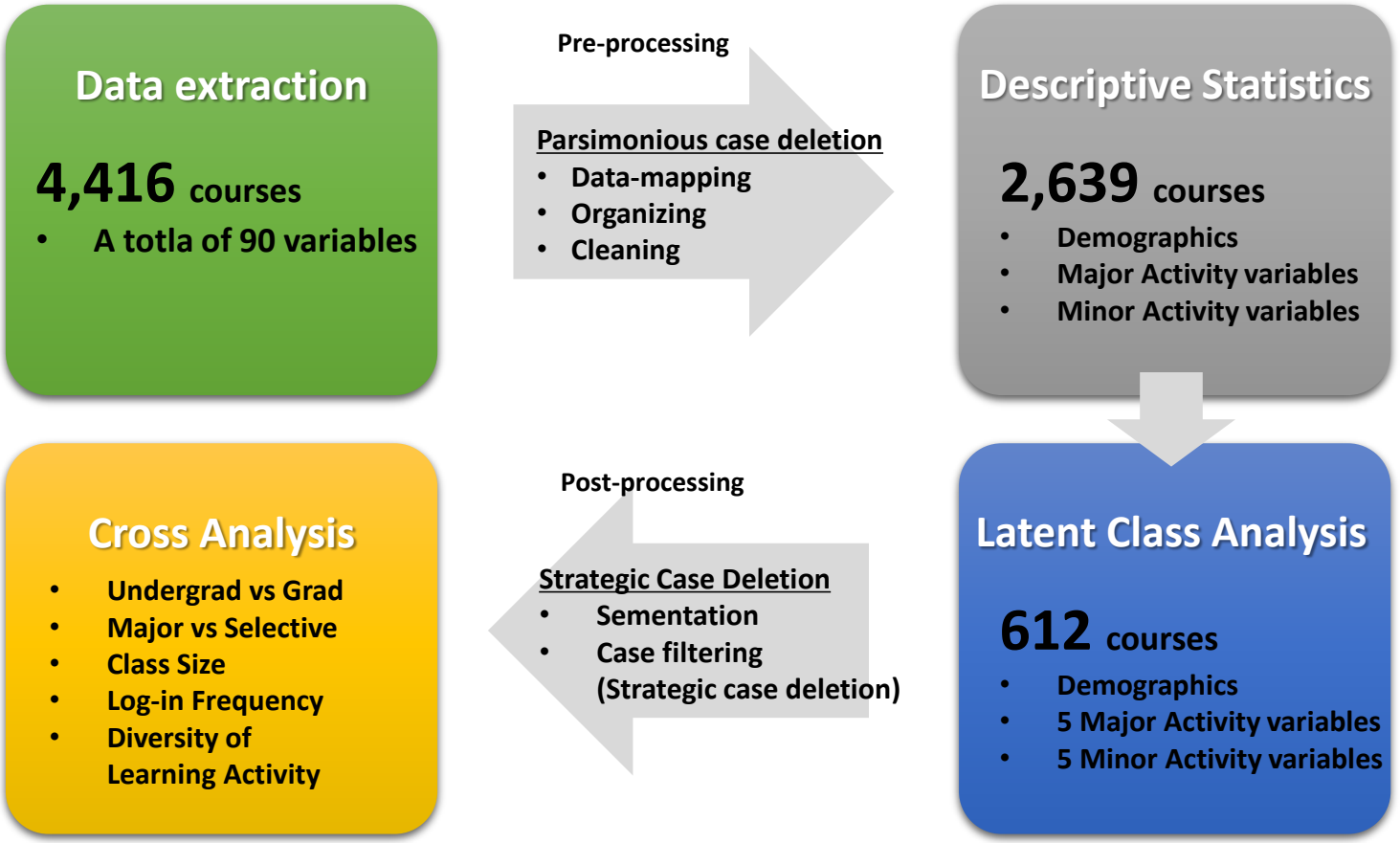
Methods

2) Descriptive Statistics (n=2,639)



Methods

3) Datamining process

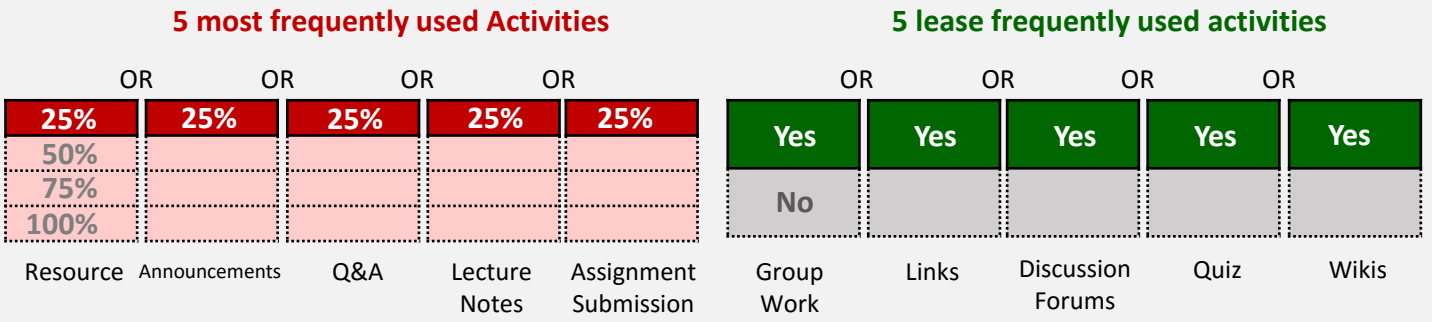


Methods

3) Datamining process

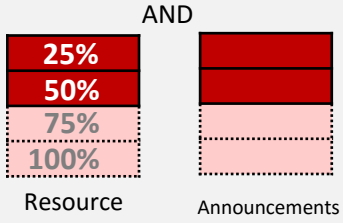
N = 2,639

Inclusion Rule 1. If at least one of the major activities belongs to **Top 25%** quartile
Inclusion Rule 2. If at least one of the minor activities belongs to **Yes** group



N = 1,804

Exclusion Rule. If both Resource and Announcements belong to **Bottom 50%** quartiles

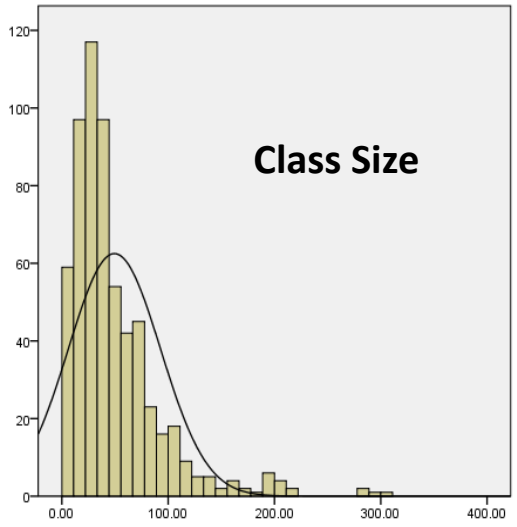


N = 612

Methods

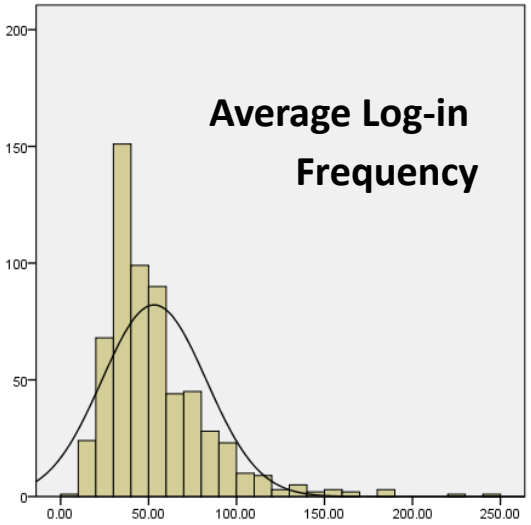
3) Datamining process → n=612

- General features : Class Size, Log-in Frequency, and Learning Activity
- Almost equally divided into 4 quartiles by using the median score



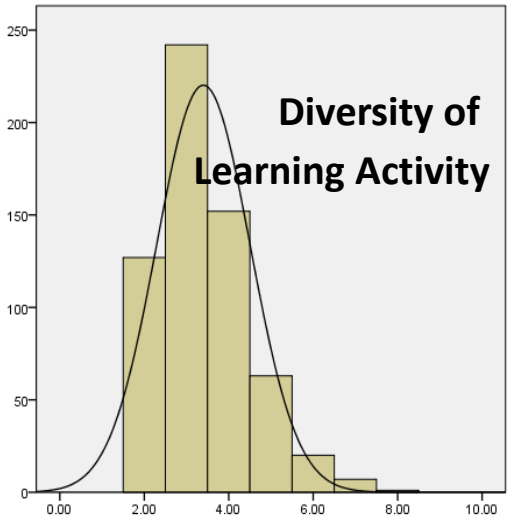
Median = 35 students

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



Median = 45 times

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



Median = 3 activities

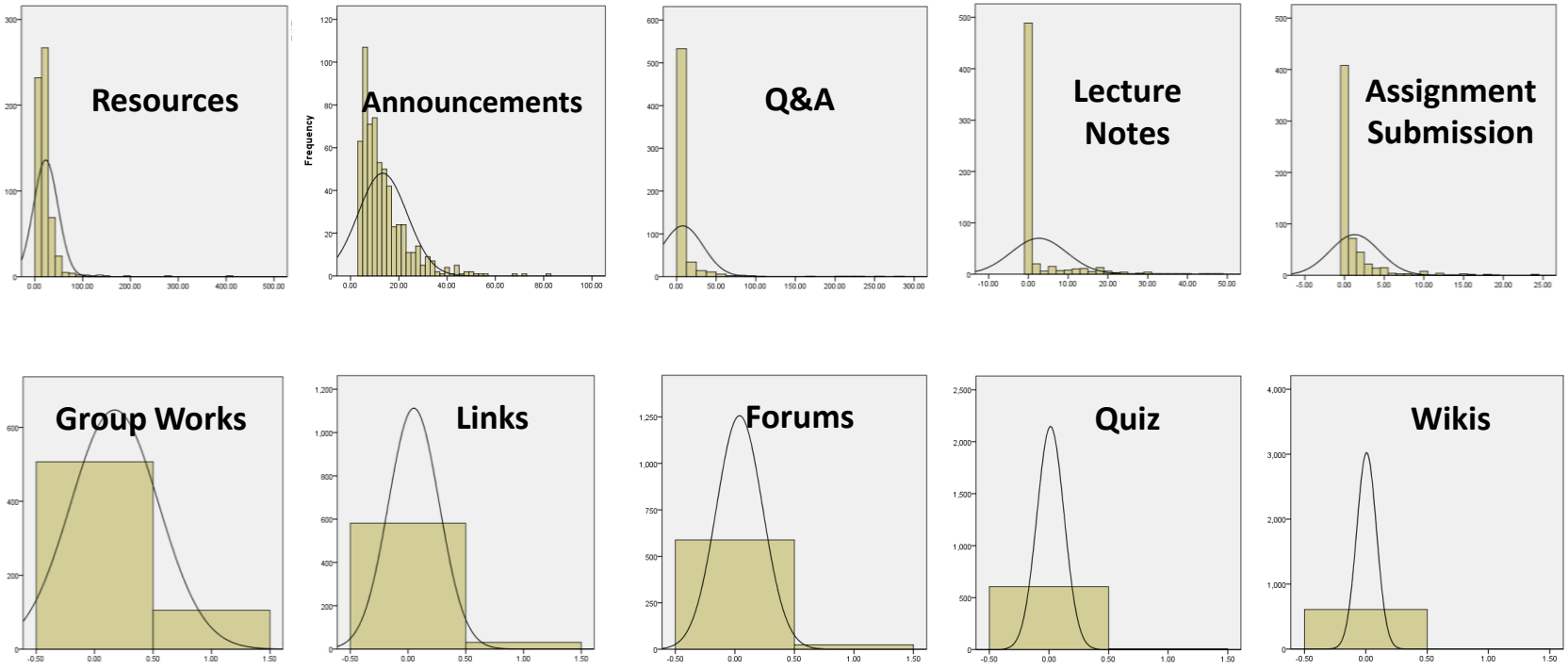
Course Type	Activity	Case
Not diverse	2	137
Little diverse	3	166
Somewhat Diverse	4	160
Very diverse	5~8	149

Methods

3) Datamining process → n=612

- Skewed & Heavy-tailed Distribution
- Categorical and Binary Data
- Conditional Independence

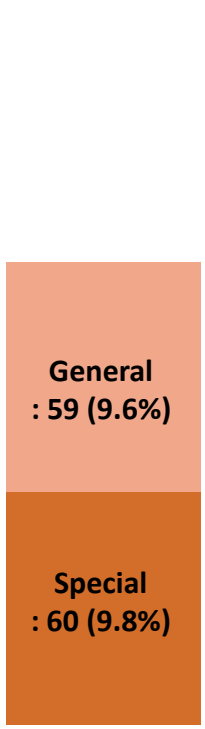
→ **Latent Class Analysis Assumptions Satisfied**



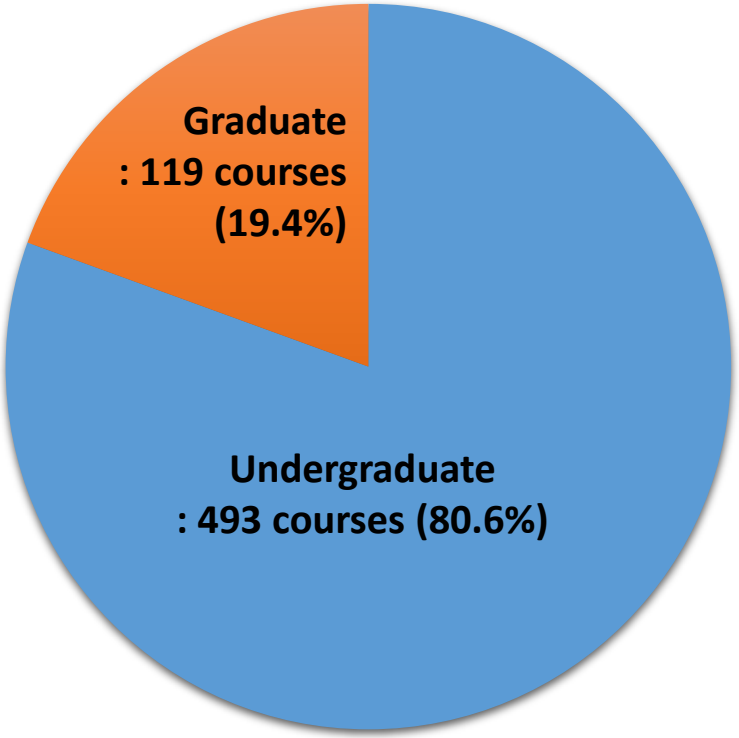
Methods

3) Datamining process → n=612

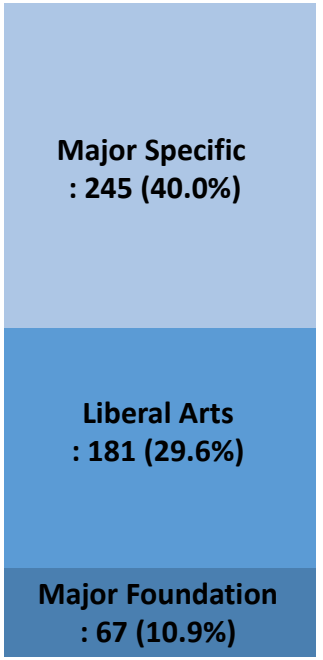
- Majority were Undergraduate-Level Major Specific & Liberal Arts
- Graduate-level courses nearly equally divided



Graduate
n = 119 (19.4%)



Total
N = 612 (100%)



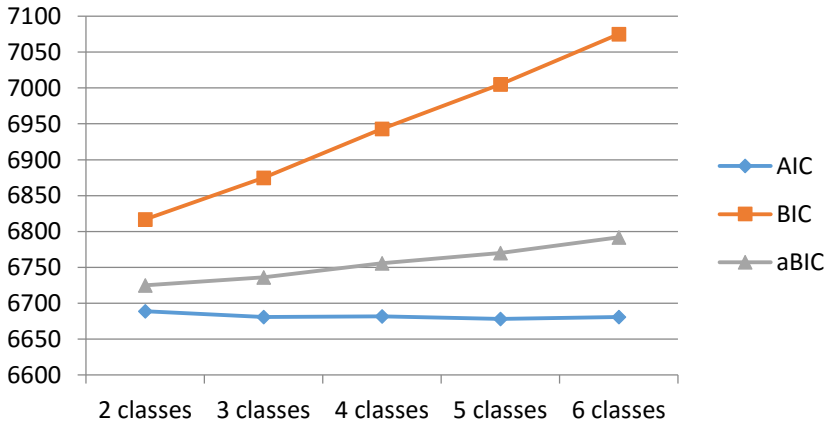
Undergraduate
n = 493 (80.6%)

Methods

4) Data Analysis (n=612)

- No single right answer to "How many latent classes are there?"
- Compared the model fit and determined 4 latent classes

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



The Smaller, The Better

- Akaike Information Criterion (AIC)
- Bayesian Information Criterion (BIC)
- the adjusted BIC (aBIC)

***p* < .05 Good Model Fit**

- Lo-Mendell-Rubin (LMR)

The Closer to 1, The Better

- Entropy

Methods

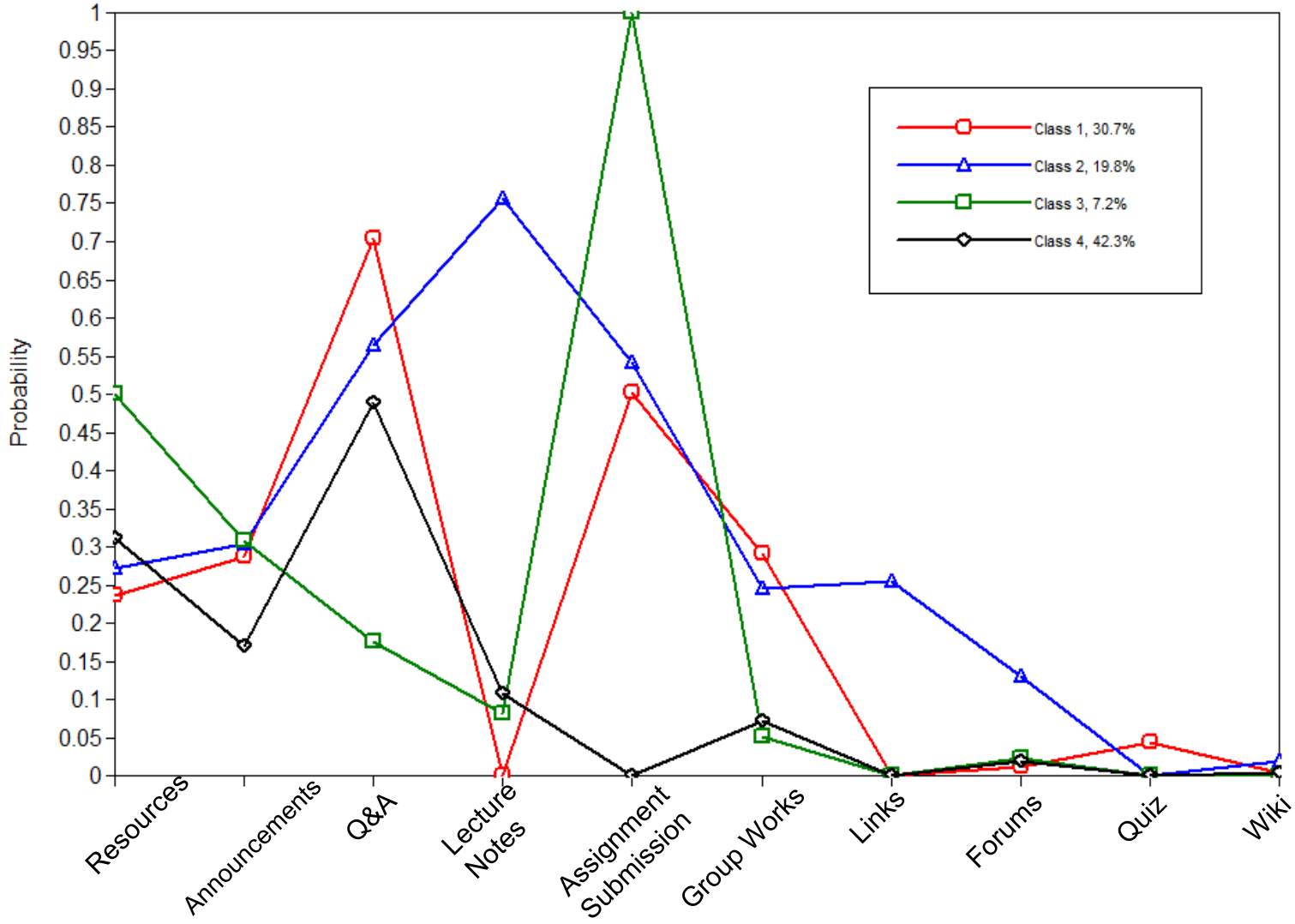
4) Data Analysis (n=612)

LCA Probability Analysis

Variables	Segmentation	Latent Class 1	Latent Class 2	Latent Class 3	Latent Class 4
Resources	1 (8~11)	.239*	.367*	.092	.118*
	2 (12~17)	236*	.272*	**.500***	.311*
	3 (18~24)	.244*	.192*	**.204***	.277*
	4 (25~401)	.281*	.169*	.203	.293*
Announcements	1 (4~5)	.051	.000	.692*	.332*
	2 (6~9)	.286*	.303*	.308	.170*
	3 (10~15)	.273*	.313*	.000	.288*
	4 (16~82)	.390*	.383*	.000	.210*
Q&A	1 (0)	.296*	.436*	.824*	.511*
	2 (1~280)	**.704***	**.564***	.176	**.489***
Lecture Notes	1 (0)	1.000	.243*	.919*	.893*
	2 (1~48)	0.000	**.757***	.081	.107*
Assignment Submission	1 (0)	.498*	.458*	.000*	1.000
	2 (1~24)	**.502***	**.542***	**1.000***	.000
Group Works	1 (0)	.709*	.755*	.949*	.929*
	2 (1~)	**.291***	**.245***	.051	.071
Links	1 (0)	1.000	.744*	1.000	1.000
	2 (1~)	0.000	**.256***	.000	.000
Forums	1 (0)	.988*	.870*	.977*	.981*
	2 (1~)	.012	**.130***	.023	.019
Quiz	1 (0)	.957*	1.000	1.000	1.000
	2 (1~)	.043	.000	.000	.000
Wiki	1 (0)	.996*	.982*	1.000	.996*
	2 (1~)	.004	.018	.000	.004

Methods

4) Data Analysis (n=612)



04. Results

1) Overall Patterns of Online Activity

LAPA
LEARNING ANALYTICS FOR PREDICTION & ACTION

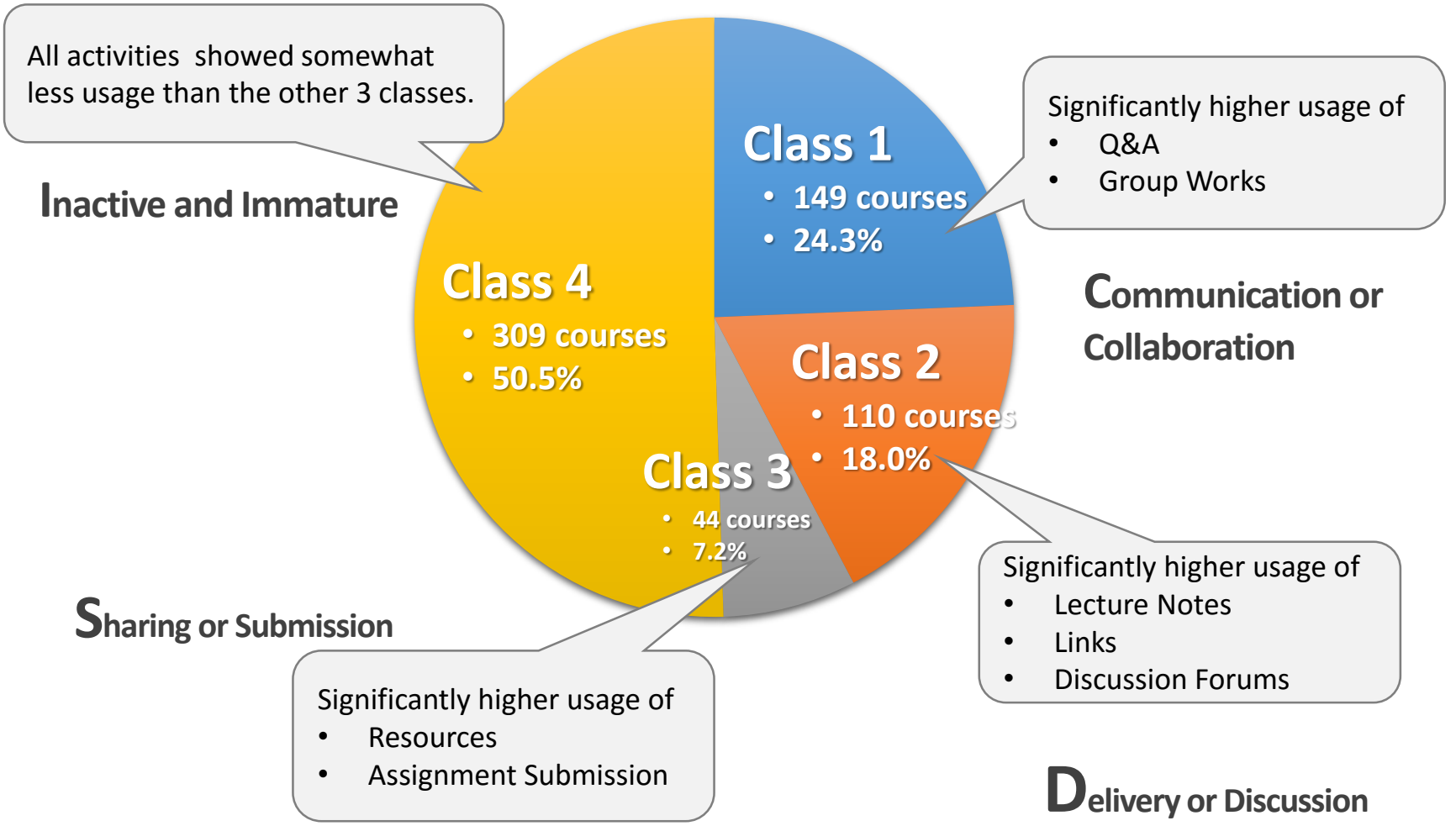
04. Results

2) Classification of BL courses

LAPA
LEARNING ANALYTICS FOR PREDICTION & ACTION

Results

2) Classification of BL courses



Results

2) Classification of blended learning courses

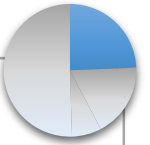
Data Post-processing

While **Class Size** was almost equally divided, both **Log-in Frequency** and **Diversity of Learning Activity** were significantly different among the 4 groups.

Variables		Class 1 (n = 149, 24.3%)	Class 2 (n = 110, 18.0%)	Class 3 (n = 44, 7.2%)	Class 4 (n = 309, 50.5%)	X^2, F
		C type	D type	S type	I type	
Class Size	Small (3~21)	27 (18.1%)	25 (22.7%)	17 (38.6%)	82 (26.5%)	12.314
	Medium (22~35)	38 (25.5%)	30 (27.3%)	12 (27.3%)	75 (24.3%)	
	M-L (36~64)	43 (28.9%)	25 (22.7%)	10 (22.7%)	73 (23.6%)	
	Large (65~301)	41 (27.5%)	30 (27.3%)	5 (11.4%)	79 (25.6%)	
Log-in Frequency	Rarely (8~32)	18 (12.1%)	7 (6.4%)	9 (20.5%)	**103 (33.3%)**	**174.176****
	Occasionally (33~44)	20 (13.4%)	8 (7.3%)	**14 (31.8%)**	**124 (40.1%)**	
	Frequently (45~66)	**58 (38.9%)**	38 (34.5%)	**13 (29.5%)**	51 (16.5%)	
	Very Frequently (67~245)	53 (35.6%)	**57 (51.8%)**	8 (18.2%)	31 (10.0%)	
Diversity of Learning Activity	Not diverse (2)	0 (0.0%)	0 (0.0%)	0 (0.0%)	127 (41.1%)	**452.479****
	Little diverse (3)	36 (24.2%)	14 (12.7%)	**30 (68.2%)**	**162 (52.4%)**	
	Somewhat diverse (4)	**86 (57.7%)**	36 (32.7%)	13 (29.5%)	17 (5.5%)	
	Very diverse (5~8)	27 (18.1%)	**60 (54.5%)**	1 (2.3%)	3 (1.0%)	

Results

2) Classification of blended learning courses Type C



[Communication or Collaboration] Class 1 = 149 courses (24.3%)

- Significantly higher usage of Q&A and Group Works
- Relatively high usage of Resources, Announcements, and Assignment Submission
- **Mostly Undergraduate-level** major-specific courses (n=66, 44.3%) from College of **Social Sciences** (n=15, 10.1%), **Education** (n=11, 7.4%), and **Liberal Arts** (n=11, 7.4%)
- Higher level of Log-in Frequency per student
- Diverse learning activities provided

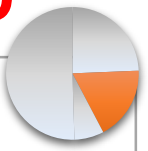
Level 1	Level 2	Level 3	n (%)
Undergraduate: 131 (87.9%)	Liberal Arts: 57 (38.3%)	**Mandatory**	**43 (28.9%)**
		Selective	14 (9.4%)
	Major Specific: 66 (44.3%)	Health Sciences	3 (2.0%)
		Business Administration	8 (5.4%)
		Engineering	8 (5.4%)
		Education	**11 (7.4%)**
		Social Sciences	**15 (10.1%)**
		Scranton	1 (0.7%)
		Pharmacy	1 (0.7%)
		Music	1 (0.7%)
		Liberal Arts	**11 (7.4%)**
		Natural Sciences	4 (2.7%)
		Art & Design	3 (2.0%)
	Major Foundation: 8 (5.4%)		
Graduate: 18 (12.1%)	General: 7 (4.7%)		
	Special: 11 (7.4%)	Business Administration	1 (0.7%)
		Education	6 (4.0%)
		International Studies	1 (0.7%)
		Design	0 (0.0%)
		Law	0 (0.0%)
		Social Welfare	2 (1.3%)
		Theology	0 (0.0%)
		Clinical Health Sciences	1 (0.7%)
		Policy Sciences	0 (0.0%)
		Translation & Interpretation	0 (0.0%)
			149 (100%)

Variables		Class 1 (n = 149, 24.3%)
Class Size	Small (3~21)	27 (18.1%)
	Medium (22~35)	38 (25.5%)
	M-L (36~64)	43 (28.9%)
	Large (65~301)	41 (27.5%)
Log-in Frequency	Rarely (8~32)	18 (12.1%)
	Occasionally (33~44)	20 (13.4%)
	Frequently (45~66)	**58 (38.9%)**
	Very Frequently (67~245)	53 (35.6%)
Diversity of Learning Activity	Not diverse (2)	0 (0.0%)
	Little diverse (3)	36 (24.2%)
	Somewhat diverse (4)	**86 (57.7%)**
	Very diverse (5~8)	27 (18.1%)

Results

2) Classification of blended learning courses

Type D



[Delivery and Discussion] Class 2 = 110 courses (18.0%)

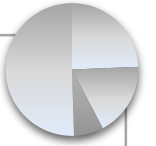
- Significantly higher usage of Lecture Notes, Links, and Discussion Forums
- Relatively high usage of Resources, Announcements, Q&A, and Assignment Submission
- Mostly Undergraduate-level major-specific courses (n=43, 39.1%) from **College of Engineering** (n=11, 10.0%) and **Social Sciences** (n=11, 10.0%)
- Higher level of Log-in Frequency per student
- Very diverse (5 or more) learning activities provided

Level 1	Level 2	Level 3	n (%)
Undergraduate: 87 (79.1%)	Liberal Arts: 34 (30.9%)	**Mandatory**	**28 (25.5%)**
		Selective	6 (5.5%)
	Major Specific: 43 (39.1%)	Health Sciences	4 (3.6%)
		Business Administration	4 (3.6%)
		Engineering	**11 (10.0%)**
		Education	7 (6.4%)
		Social Sciences	**11 (10.0%)**
		Scranton	0 (0.0%)
		Pharmacy	0 (0.0%)
		Music	1 (0.9%)
		Liberal Arts	3 (2.7%)
		Natural Sciences	1 (0.9%)
		Art & Design	1 (0.9%)
	Major Foundation: 10 (9.1%)		
Graduate: 23 (20.9%)	General: 15 (13.6%)		
	Special: 8 (7.3%)	Business Administration	0 (0.0%)
		Education	4 (3.6%)
		International Studies	0 (0.0%)
		Design	0 (0.0%)
		Law	1 (0.9%)
		Social Welfare	0 (0.0%)
		Theology	0 (0.0%)
		Clinical Health Sciences	2 (1.8%)
		Policy Sciences	1 (0.9%)
		Translation & Interpretation	0 (0.0%)
			110 (100%)

Variables		Class 2 (n = 110, 18.0%)
Class Size	Small (3~21)	25 (22.7%)
	Medium (22~35)	30 (27.3%)
	M-L (36~64)	25 (22.7%)
	Large (65~301)	30 (27.3%)
Log-in Frequency	Rarely (8~32)	7 (6.4%)
	Occasionally (33~44)	8 (7.3%)
	Frequently (45~66)	38 (34.5%)
	Very Frequently (67~245)	**57 (51.8%)**
Diversity of Learning Activity	Not diverse (2)	0 (0.0%)
	Little diverse (3)	14 (12.7%)
	Somewhat diverse (4)	36 (32.7%)
	Very diverse (5~8)	**60 (54.5%)**

Results

2) Classification of blended learning courses Type S



[Sharing or Submission] Class 3 = 44 courses (7.2%)

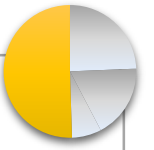
- Significantly higher usage of Resources and Assignment Submission
- No usage of any other functions
- Mostly Undergraduate-level major-specific courses (n=26, 59.1%) from **College of Education** (n=8, 18.2%) and **Business Administration** (n=5, 11.4%)
- Relatively low level of Log-in Frequency per student
- Learning activities were less diverse

Level 1	Level 2	Level 3	n (%)
Undergraduate: 32 (72.7%)	Liberal Arts: 3 (6.8%)	Mandatory	2 (4.5%)
		Selective	1 (2.3%)
	Major Specific: 26 (59.1%)	Health Sciences	1 (2.3%)
		Business Administration	5 (11.4%)
		Engineering	1 (2.3%)
		Education	**8 (18.2%)**
		Social Sciences	4 (9.1%)
		Scranton	2 (4.5%)
		Pharmacy	2 (4.5%)
		Music	0 (0.0%)
		Liberal Arts	1 (2.3%)
		Natural Sciences	2 (4.5%)
		Art & Design	0 (0.0%)
	Major Foundation: 3 (6.8%)		
Graduate: 12 (27.3%)	General: 5 (11.4%)		
	Special: 7 (15.9%)	Business Administration	0 (0.0%)
		Education	3 (6.8%)
		International Studies	1 (2.3%)
		Design	0 (0.0%)
		Law	0 (0.0%)
		Social Welfare	0 (0.0%)
		Theology	0 (0.0%)
		Clinical Health Sciences	3 (6.8%)
		Policy Sciences	0 (0.0%)
		Translation & Interpretation	0 (0.0%)
			44 (100%)

Variables		Class 3 (n = 44, 7.2%)
Class Size	Small (3~21)	17 (38.6%)
	Medium (22~35)	12 (27.3%)
	M-L (36~64)	10 (22.7%)
	Large (65~301)	5 (11.4%)
Log-in Frequency	Rarely (8~32)	**9 (20.5%)**
	Occasionally (33~44)	**14 (31.8%)**
	Frequently (45~66)	**13 (29.5%)**
	Very Frequently (67~245)	8 (18.2%)
Diversity of Learning Activity	Not diverse (2)	0 (0.0%)
	Little diverse (3)	**30 (68.2%)**
	Somewhat diverse (4)	13 (29.5%)
	Very diverse (5~8)	1 (2.3%)

Results

2) Classification of blended learning courses Type I



[Inactive or Immature] Class 4 = 309 courses (50.5%)

- **Except resources and assignment submission, there was almost no usage of online lea rning activities**
- Mostly undergraduate-level major-specific courses (n=110, 35.6%) from College of Biz Administration, Social Sciences, Liberal Arts, Education, and Natural Sciences

Level 1	Level 2	Level 3	n (%)
Undergraduate: 243 (78.6%)	Liberal Arts: 87 (28.2%)	Mandatory	68 (22.0%)
		Selective	19 (6.1%)
	Major Specific: 110 (35.6%)	Health Sciences	5 (1.6%)
		Business Administration	**23 (7.4%)**
		Engineering	9 (2.9%)
		Education	**16 (5.2%)**
		Social Sciences	**20 (6.5%)**
		Scranton	0 (0.0%)
		Pharmacy	5 (1.6%)
		Music	2 (0.6%)
		Liberal Arts	**17 (5.5%)**
		Natural Sciences	**12 (3.9%)**
		Art & Design	1 (0.3%)
	Major Foundation: 46 (14.9%)		
Graduate: 66 (21.4%)	General: 32 (10.4%)		
	Special: 34 (11.0%)	Business Administration	4 (1.3%)
		Education	13 (4.2%)
		International Studies	3 (1.0%)
		Design	1 (0.3%)
		Law	1 (0.3%)
		Social Welfare	0 (0.0%)
		Theology	2 (0.6%)
		Clinical Health Sciences	2 (0.6%)
		Policy Sciences	2 (0.6%)
		Translation & Interpretation	6 (1.9%)
			309 (100%)

Variables		Class 4 (n = 309, 50.5%)
Class Size	Small (3~21)	82 (26.5%)
	Medium (22~35)	75 (24.3%)
	M-L (36~64)	73 (23.6%)
	Large (65~301)	79 (25.6%)
Log-in Frequency	Rarely (8~32)	**103 (33.3%)**
	Occasionally (33~44)	**124 (40.1%)**
	Frequently (45~66)	51 (16.5%)
	Very Frequently (67~245)	31 (10.0%)
Diversity of Learning Activity	Not diverse (2)	**127 (41.1%)**
	Little diverse (3)	**162 (52.4%)**
	Somewhat diverse (4)	17 (5.5%)
	Very diverse (5~8)	3 (1.0%)

05. Conclusion and Discussion

LAPA
LEARNING ANALYTICS FOR PREDICTION & ACTION

05 Conclusion and Discussions

1

Online behavior data tracked from LMS indicated the adoption level and patterns of blended learning implementation effectively.

→ Showing "adoption and early implementation" (level 2) in the blended learning adoption spectrum (Grahamet al., 2013).

→ 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.

2

Latent Class Analysis clustered 612 courses into **four types with C-D-S-I.**

→ **C(Communication or Collaboration**: 24.3%
→ D**(Delivery or Discussion):** 18%
→ **S(Sharing or Submission): 7.2%**
→ **I (Inactive or Immature): 50%**

Active

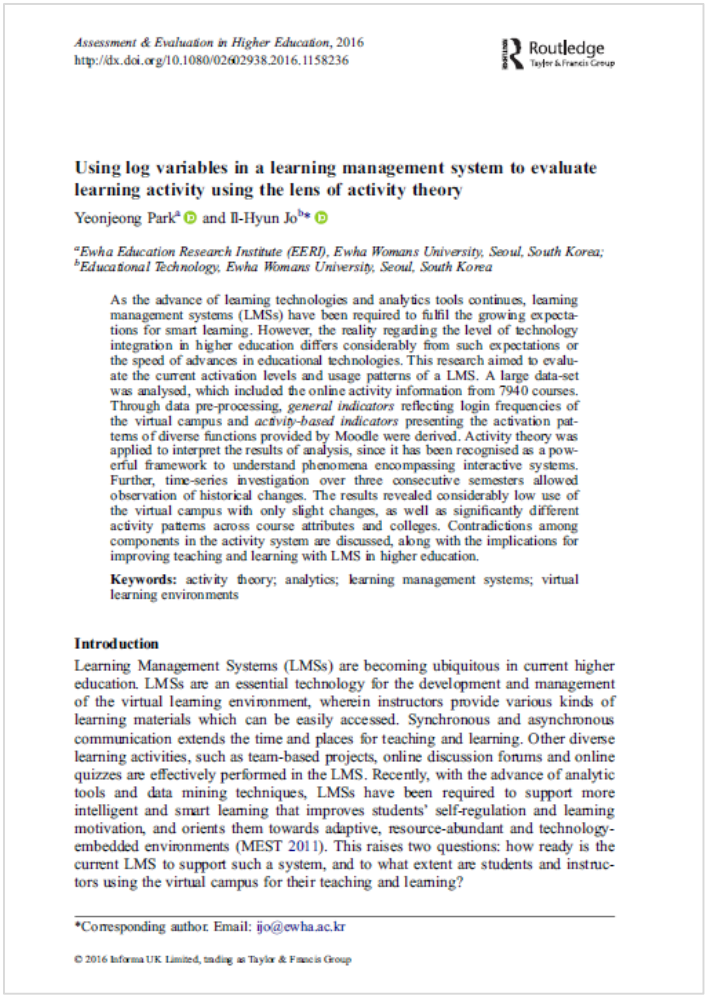
Passive engagement of online technology

3

The **data-driven approach in institution-level** helps academic leaders and staffs monitor the whole status of BL and to invest developing LMS in a strategic way.

→ Requiring to develop a model that fits to the ecological nature of diverse classes, and the advanced technologies that support diverse pedagogical models.

06 For the detailed contents ...



Assessment & Evaluation in Higher Education, 2016
http://dx.doi.org/10.1080/02602938.2016.1158236

Routledge
Taylor & Francis Group

Using log variables in a learning management system to evaluate learning activity using the lens of activity theory

Yeonjeong Park[a] and Il-Hyun Jo[b]*

[a]*Ewha Education Research Institute (EERI), Ewha Womans University, Seoul, South Korea;*
[b]*Educational Technology, Ewha Womans University, Seoul, South Korea*

As the advance of learning technologies and analytics tools continues, learning management systems (LMSs) have been required to fulfil the growing expectations for smart learning. However, the reality regarding the level of technology integration in higher education differs considerably from such expectations or the speed of advances in educational technologies. This research aimed to evaluate the current activation levels and usage patterns of a LMS. A large data-set was analysed, which included the online activity information from 7940 courses. Through data pre-processing, *general indicators* reflecting login frequencies of the virtual campus and *activity-based indicators* presenting the activation patterns of diverse functions provided by Moodle were derived. Activity theory was applied to interpret the results of analysis, since it has been recognised as a powerful framework to understand phenomena encompassing interactive systems. Further, time-series investigation over three consecutive semesters allowed observation of historical changes. The results revealed considerably low use of the virtual campus with only slight changes, as well as significantly different activity patterns across course attributes and colleges. Contradictions among components in the activity system are discussed, along with the implications for improving teaching and learning with LMS in higher education.

Keywords: activity theory; analytics; learning management systems; virtual learning environments

Introduction

Learning Management Systems (LMSs) are becoming ubiquitous in current higher education. LMSs are an essential technology for the development and management of the virtual learning environment, wherein instructors provide various kinds of learning materials which can be easily accessed. Synchronous and asynchronous communication extends the time and places for teaching and learning. Other diverse learning activities, such as team-based projects, online discussion forums and online quizzes are effectively performed in the LMS. Recently, with the advance of analytic tools and data mining techniques, LMSs have been required to support more intelligent and smart learning that improves students' self-regulation and learning motivation, and orients them towards adaptive, resource-abundant and technology-embedded environments (MEST 2011). This raises two questions: how ready is the current LMS to support such a system, and to what extent are students and instructors using the virtual campus for their teaching and learning?

*Corresponding author. Email: ijo@ewha.ac.kr

Park, Y., & Jo, I. (2016). Using log variables in a learning management system to evaluate learning activity using the lens of activity theory. Assessment & Evaluation in Higher Education, 1-17.

Internet and Higher Education 29 (2016) 1–11

Contents lists available at ScienceDirect

Internet and Higher Education

ELSEVIER

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. 2013S1A5A2A03044310).

* 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/

Park, Y., Yu, J., & Jo, I. (2016). Clustering blended learning courses by online behavior data: A case study in a Korean higher education institute. *The Internet and Higher Education, 29*(2016), 1-11.

Thank you

Yeonjeong Park ypark@honam.ac.kr / chocolike@gmail.com