Learning Management System Analytics to Examine the Behavior of Students in High Enrollment STEM Courses During the Transition to Online Instruction

Rob Elliott, Ed.D.
Computer and Information Technology
Indiana University Purdue University Indianapolis (IUPUI)
Indianapolis, Indiana
elliott@iupui.edu

Xiao Luo, Ph.D.
Computer and Information Technology
Indiana University Purdue University Indianapolis (IUPUI)
Indianapolis, Indiana
luo25@iupui.edu

***Abstract*—The emergence of the COVID-19 pandemic resulted in the transition to near-total online instruction in early 2020. Several studies surveyed students about the impact of the pandemic on their behavior and engagement with their education; however, those studies may not include analysis of actual student behaviors. The field of learning analytics allows researchers to examine the records made while students interact with the educational technology tools that are commonly used to facilitate instruction in Institutes of Higher Education (IHEs). In most universities, delivery of many instructional materials is conducted via a Learning Management System (LMS). In this study, we describe our process to examine student interactions within the LMS to discover any measurable changes to student behavior during the pandemic.**

We examined the usage logs of the Canvas LMS at a large university in the midwestern US to examine the behavior of students' interactions with high-enrollment STEM courses in two semesters: one prior to and one during the pandemic. The log data was integrated with student demographic data so that the LMS behavior of subsets of students can be compared. Machine learning algorithms including clustering models and association rule mining were applied on the data. The results of this study demonstrate that the students' behaviors did change in the transition to online instruction. Students had more frequent sessions in the LMS on both computers and mobile devices, although the duration of their mobile sessions was shorter after courses were moved online. Further, students in historically underrepresented groups in STEM fields were found to use their mobile devices more frequently for academic work. The information uncovered in this study can be used to inform future instructional design practices with the LMS to promote an equitable experience for all students.

***Index Terms*—Educational technology, STEM, Mobile learning, Learning management systems**

I. Introduction

The spread of COVID-19 in early 2020 resulted in the physical closure of most academic facilities around the world and forced most institutes of higher education (IHEs) to an exclusively online academic model [1]. The transition to online learning launched a flurry of studies to gauge its impact on students. Several studies focused on students' personal technology and access to digital course materials [2]–[4]. Others asked students to describe how the pandemic affected their engagement and relationship with school [5], [6].

It is understandable that the pandemic's impact affected students' ability to complete their coursework. However, simple surveys may not provide a complete picture of how students' behaviors were affected by the transition to online learning. The inability to access campus computing resources was a challenge for some students, particularly those who already face socioeconomic challenges in attending college [7]. Thus, some students were forced to find alternative resources or use the resources they had at their disposal differently in order to engage in their academic work [4].

This study describes the application of learning analytics to investigate student behavior with learning technology in an effort to quantitatively describe how technology use may have been impacted by the pandemic. Our intent is twofold: first, to describe the process used to analyze students' digital footprints, and second, to use the data obtained from learning technology to answer research questions. The research questions at hand are:

1) How did the transition to online learning change students' use of their personal technological devices for academic work at a large public institution in the US?
2) Did the behaviors of any subgroups of students change when compared to the student body as a whole at a large public institution in the US?

Specifically, this study compares student behavior relative to the devices used to access the Canvas Learning Management System (LMS) at a large multi-campus public university in the midwestern United States. This site was selected for this study because the transition to online learning, although abrupt, did not require the adoption of additional learning technology when coursework was moved online. The Canvas LMS had been heavily used by the institution in the years prior to the pandemic. Therefore, we had access to a robust backlog of student behavior data from academic terms prior to the pandemic that we could use as a reference to compare student behavior during the impacted semesters.

The process used to obtain and analyze the data from the Canvas LMS is applicable to any IHE that uses the specific Canvas product, and should be generalizable to any IHE that uses any of the prominent LMS technologies. The educational application of this research is to describe how students interact with digital course content while using their personal technological devices. This information can provide valuable insight to instructors, instructional designers, and the decision-makers at IHEs who select the learning technologies used by faculty and students. In summary, this study uses the automated data generated through interactions with learning technology to provide a quantitative description of students' behavioral differences based on the transition to online learning.

II. Literature Review

The intersection of higher education and widespread internet access has transformed the tools used for teaching. The past few decades have shown an increasing trend toward the digitalization of course materials, including electronic textbooks and interactive content [8]–[10]. This phenomenon may have somewhat cushioned the transition to online learning during the COVID-19 pandemic [11]. Similarly, students were generally (although not universally) well-equipped with the personal technology that allowed them to continue their studies in higher education, particularly in the United States [1], [4]. Several studies have demonstrated that students use a variety of devices to access course materials and perform academic work in higher education [12], [13].

LMS implementation is nearly universal at IHEs in the United States [14] and were used by 88% of faculty even prior to the pandemic. Instructors primarily use the LMS for the dissemination of academic content to students [15], which can include texts, handouts, instructional videos, discussion boards, assignments, and quizzes or exams [16]. An LMS facilitates the transfer of digital course materials between a faculty and student (and vice-versa) with the intent of providing more control to the learners by affording them more flexibility in the time and space where they engage with the content [17].

LMSs and other platforms allow the collection of both targeted and passive data about the particular activities performed within the software and the user who is performing the activities. For example, as students move through the Canvas LMS, each request made of the web server is recorded. These requests include information about the user, course, device used, activity performed, and the time of the request itself [18]. The collection of requests is stored in a log file, which is ripe for data analysis [19]. This raw data is the foundation of most projects that describe learning analytics. Mella-Norambuena, et al, [13], for example, used Canvas log data to examine the behavior of students with regards to their mobile devices in Chile.

At a minimum, learning analytics can be defined as "...the application of (Big Data) techniques to improve learning" [20, p. 684]. More robust definitions note that a primary goal of the examination of learning analytics is "...to better understand learner behaviors and context (both digital and analog) to improve learning outcomes and to increase institutional efficiency and effectiveness" [21, p. 144].

Further, LMSs are often delineated by a specific course and even more granularly by course sections, which are offered in a variety of modalities by IHEs. In the United States, virtually every IHE delivered their coursework online after the emergence of COVID-19 during March 2020, and over 72% of courses were delivered in an online or distance modality during the Fall 2020 semester [1].

Few formal definitions for the modality of a course at an IHE exist. Many higher education instructors argue that the transition from in-person to online instruction during March 2020 should be recorded as "emergency remote teaching" rather than "online learning" [22]. Beyond that, there exists little agreement about the nuanced definition of what constitutes an "online course". Hodges, et. al., [22] provides a good overview of the terminology and factors that lead to the different course modality definitions.

III. Methods

The site for this study is a multi-campus IHE in the midwestern United States that uses the traditional North American Fall/Spring semester schedule. The study site's transition to online learning occurred in March 2020 — the middle of the Spring semester. Therefore, to analyze the impact of the transition, it was decided to compare student behavior from the Fall 2019 semester (the last full academic semester prior to the impact of the pandemic) to the Fall 2020 semester (the first full academic semester that dealt with the pandemic's complications). The decision was made to examine students' behavior in the Canvas LMS, as that was the most commonly-implemented learning technology at the university and 100% of courses offered at the study site have a course site established in the LMS, whether or not it is used by the instructor or made available to students.

A. Build Study Data

To examine student behavior in the LMS aligned with their demographic information in the context of specific courses required the development of an events database where each student interaction could be examined. Doing so required the consolidation of data from four separate systems: a database of courses offered, sources for student enrollment and demographic information, and the Canvas LMS data logs for the courses included in the study. A simplified overview of the data sources used is included in Figure 1.

The focus of this study was narrowed to student behavior in high enrollment STEM courses. To begin, we filtered all of the courses offered at the study site in the target semesters. All of the programs offered at the study site are assigned to a Classification of Instructional Program (CIP) code, as defined by the National Center for Education Statistics [23]. The authors analyzed courses offered by any program whose CIP code is classified as a "STEM field" by the Department of Homeland Security (DHS) Optional Practical Training (OPT)

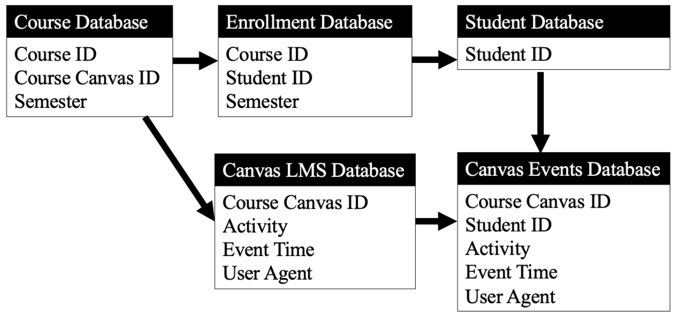

Fig. 1. Schema of Data Sources

program [24]. Therefore, only courses offered by a program that meets the DHS OPT definition of STEM are included in the analysis.

Once STEM programs were identified, data about each of the courses offered by those programs was extracted from the university's course scheduling system. The specific data requested for each course included:

- Semester offered
- Course Canvas ID
- Program offering course
- Course level (100-800, where 100-400 level courses are undergraduate courses)
- Modality
- Enrollment after first week

The *Course Canvas ID* field refers to the unique identifier for this course in the Canvas LMS, which allows us to connect information about the course to the LMS records. To ensure that only courses with high enrollment were included, only courses that had 60 or more students after the first week of the semester were included in our analysis.

We then requested demographic information about each of the students enrolled in any of the courses included in the study. This data was retrieved from the university's student information system and included the following features:

- Student ID (encrypted for student privacy)
- Age
- Gender
- Ethnicity
- School/College of enrollment
- Major

Finally, we extracted the raw log files from the Canvas LMS for courses that met the criteria for the study in the two target semesters. This data is stored as a JSON object as described in the Canvas API [18], but was transformed into a relational database with the following features extracted into columns:

- Student ID (encrypted for student privacy, but aligned with the Student database)
- Course Canvas ID
- Event Timestamp
- Request URL
  - Asset Type
  - Asset Subtype
- User Agent
  - Device Family
  - Operating System
  - Software Type (e.g., app or website)

The *Request URL* field was parsed to extract the *Asset Type* and *Asset Subtype* information from the path requested by the user. The *User Agent* field was similarly parsed and provided the detail for the *Device Family*, *Operating System*, and *Software Type* fields [25]. The data from the user agent thus revealed the information about the students' specific device that was used to interact with the LMS. Further, the mobile applications developed by Canvas contain unique user agents that allow us to determine the operating system and device for interactions via mobile devices [26].

B. Derive Learning Behavioral Elements

In this study, we use the term **event** to describe a specific request made by a student to the LMS. Analyzing the events can tell us what action the student was performing at that point in time. As a student proceeds through a sequence of events in the course of their interaction with the LMS, those related events are grouped into a session. We chose to define a **session** as the series of events performed by one student on a single device family within a single course in the LMS. A session is considered concluded when a student moves into another course in the LMS or after 25 minutes of inactivity. A **concurrent session** is a session where both a mobile device *and* a computer are in use simultaneously for at least a portion of the session.

This study also analyzes the device family used by the student to engage with the LMS. The user agent provides information about the operating system and device type in use, but does not provide in-depth information about the devices themselves (e.g., laptop or desktop). To simplify the analysis, we chose to use the operating system of the device to divide the sample into two categories of device family: **mobile devices** are those that run a mobile operating system (such as iOS, iPadOS, and Android), whereas **computers** run a full version of an operating system (MacOS, Linux, or Windows). Chromebooks [27] are used by a small but increasing number of students in higher education [28] and likely should be considered a third device family given their limited operating system and laptop-like physical features. However, the user agent recorded by requests from ChromeOS is indistinguishable from the Chrome browser used on a laptop or desktop. Therefore, for the purposes of this study, Chromebook requests are included in the computer device family.

C. Generate Behavioral Clusters

Beyond reviewing the student behaviors in aggregate, another objective of this study is to discover demographic subgroups of students whose behaviors may have changed more than the average. Therefore, we applied the k-means algorithm to generate clusters based on the learning behavioral clusters based on the following elements to describe students learning behaviors:

- Total number of sessions on a computer
- Total number of sessions on a mobile device
- Total number of concurrent sessions
- Average duration of sessions on a computer
- Average duration of sessions on a mobile device

The k-means algorithm has been used in various application domains. K-means is rather easy to implement and apply even on large and high dimensional data sets. The algorithm assigns the data to one of the clusters. The objective is to minimize the sum of the distances of the points within the cluster to the cluster center. Given $N = \{x_1, x_2, \ldots, x_i\}$ is a set of d-dimensional input data, and $\mu = \{\mu_1, \mu_2, \ldots, \mu_k\}$ is a set of initialized k centers with d-dimension, the algorithm can be summarized as following:

- Assignment cluster center: Assign each input data x_i to the cluster μ_j to which cluster center it is the minimum of all the cluster centers

$$\mu_j = x_i : ||x_i - \mu_j|| \leq ||x_i - \mu_i||, \forall i, 1 \leq i \leq k \quad (1)$$

- Update cluster center: Set the new center of each cluster to the mean of all data points belongs to that cluster

$$\mu_i = \frac{1}{|\mu|} \sum_{j \in u_i} x_j, \forall i \quad (2)$$

- Repeat the previous two steps until convergence

The silhouette value [29] is used to measure the clustering results of k-means to determine the optimal number of clusters. Silhouette value measures the separation of the resulting clusters. In other words, it measures similarity of an instance is to its own cluster compared to other clusters. In this research, we experiment with k values from 2 to 35 for k-means. The optimal k is chosen when the silhouette value average of all instances is high.

D. Interpret the Associations Between Behaviors and Student

In order to investigate the associations between the student learning behaviors and demographics and majors, the association rule mining (ARM) was applied to each behavioral cluster. ARM algorithm was originally designed by Agrawal, et al., [30] to find items that occur simultaneously and frequently in database transactions. In this study, we used it to find the associations between students features in the behavioral clusters. Given a set of attribute values $V = \{v_1, v_2, \ldots, v_m\}$ and a set of students P, each student p_i is represented by its attribute values $p_i = \{v_{i1}, v_{i2}, ..., v_{im}\}$ and $v_{ij} \in V$, ARM finds a set of rules in the form of Equation 3, where $X, Y \subseteq V$ are sets of attribute values, and $X \cap Y = \emptyset$. The rule reflects the likelihood of the co-occurrences of X and Y when X occurs in a record.

$$X \rightarrow Y \quad (3)$$

The importance of a rule is defined by the support rate (S), the confidence rate (C) and the lift value (L) [30], as shown in Equations 4, 5 and 6. The support rate and confidence rate indicate the rule frequency and the frequency of the attribute values Y appearing in transactions containing attribute values X, respectively. The confidence rate reflects the level of reliability of a rule. The lift value measures the frequency of X and Y appear together while considering the popularity of Y in the data set. The lift value reflects the interestingness of the rule; that is, the higher the lift value is, the more interesting the rule is.

$$S(X \rightarrow Y) = \frac{|X \cup Y|}{|PT|} \quad (4)$$

$$C(X \rightarrow Y) = \frac{|X \cup Y|}{|X|} \quad (5)$$

$$L(X \rightarrow Y) = \frac{S(X \rightarrow Y)}{S(X)S(Y)} \quad (6)$$

Once the clusters were discovered, we conducted ARM on the two populations of students (Fall 2019 and Fall 2020) by analyzing the following demographic features:

- Gender
- School of Major
- Age Group
- Race/Ethnicity

IV. Results

A. Statistics of the Study Data

We found a total of 623 courses (inclusive of multiple sections, such as a Monday/Wednesday class and a Tuesday/Thursday class of the same course) in the Fall 2019 and Fall 2020 semesters, as described in Table I. At the study site, the term ”hybrid” is used to describe a course that includes students in the physical classroom with students joining via videoconferencing.

TABLE I
Course Details

Category	***Fall 2019***	***Fall 2020***	***Total***
In-Person	291	5	296
Asynchronous Online	15	104	119
Synchronous Online	4	27	31
Hybrid	11	167	178
Total	321	302	623

1) Student Details: In the target semesters, 27,224 unique students engaged with the LMS at least once in a high enrollment STEM course. 13,168 (48.4%) students appear in the dataset twice, meaning they took at least one high-enrollment STEM course in both Fall 2019 and Fall 2020.

Traditional statistical analysis shows that students exhibited differences in both the number and duration of LMS sessions between the Fall 2019 and Fall 2020 semesters. Students used their computers to access the LMS more frequently in Fall 2020 than in Fall 2019. In Fall 2019 students accessed their courses in the LMS an average of 203 times throughout the course of the semester compared to an average of 308 times

TABLE II
STUDENT DETAILS

	Fall 2019	*Fall 2020*
Total Number of Students	24,354	16,067
Used a Computer	24,311 (99.8%)	15,949 (99.7%)
Used a Mobile Device at Least Once	23,248 (95.5%)	15,322 (95.7%)
Used a Mobile Device >5 Times	21,000 (86.2%)	14,089 (88.0%)
Did Not Use a Computer	43 (0.2%)	54 (0.3%)
Did Not Use a Mobile Device	1,106 (4.5%)	681 (4.3%)

in the Fall 2020 semester. The number of computer-based sessions per student did not follow a normal distribution, so a Mann-Whitney test was used to compare the number of computer-based sessions between the semesters. The test indicated that students had a higher median number of computer-based sessions in Fall 2020 (Median=165) than in the Fall 2019 semester (Median=119), U=151669257.5, $p<.001$ with a moderate effect size. Similarly, another Mann-Whitney test supports the finding that students on average also used their mobile devices to access the LMS more frequently in Fall 2020 (Median=66) than in Fall 2019 (Median=40), U=137616303.0, p=0.0 with a moderate effect size.

TABLE III
SESSION & EVENT DETAILS

Category	*Fall 2019*	*Fall 2020*
Number of Sessions	**4,951,051**	**4,959,324**
Computer	3,503,187 (70.1%)	3,313,768 (66.9%)
Mobile	1,446,376 (29.2%)	1,644,075 (33.2%)
Number of Events	**20,015,456**	**28,684,132**
Computer	16,060,202 (80.2%)	21,531,046 (75.1%)
Mobile	3,955,254 (19.8%)	7,153,086 (24.1%)

2) Session Details: 48,702,989 events occurred in 9,910,375 sessions total for all courses in both semesters of the study. There was an increase in the number of mobile sessions between Fall 2019 and Fall 2020; 29.2% of sessions were conducted on a mobile device in Fall 2019, whereas 33.2% of sessions were conducted on a mobile device in Fall 2020. In conjunction with the increase number of sessions on mobile devices in Fall 2020, students also engaged more with the LMS using those devices. 24.1% of Fall 2020 events conducted in the LMS were done so on a mobile device, an increase from 19.8% in Fall 2019.

Students used their computers for longer sessions in Fall 2020 than they did in Fall 2019. A Mann-Whitney test indicated that sessions on computers were longer in the Fall 2020 semester (Median=5.7 minutes) than in the Fall 2019 semester (Median=4.5 minutes), U = 134016355.5, p=0.0 with a moderate effect size. And although students used their mobile devices more frequently in Fall 2020 when compared to Fall 2019 they did so for shorter durations, as confirmed by a Mann-Whitney test measuring the median length of a session on a mobile device in Fall 2019 (Median=0.97) and Fall 2020 (Median=0.8), U = 200970487.0, $p<0.001$ with a large effect size.

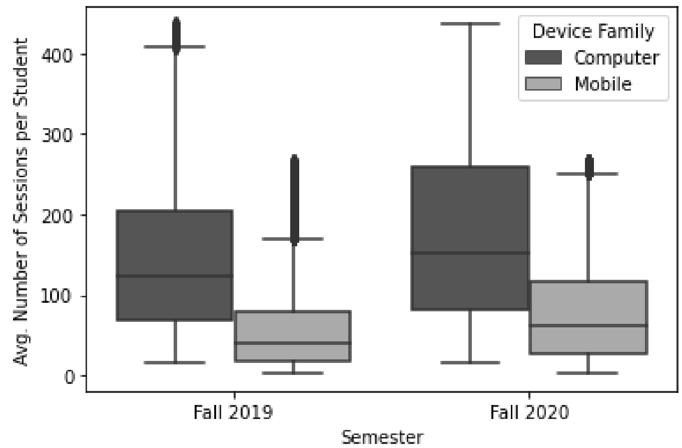

Fig. 2. Comparison of Mean Number of Sessions per Device Family (Trimmed 5%-95%)

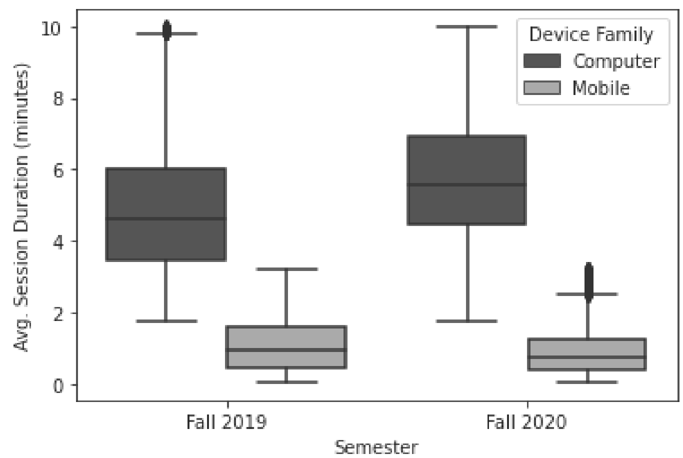

Fig. 3. Comparison of Mean Session Time per Device Family (Trimmed 5%-95%)

B. Behavior Clusters

We implemented the k-means algorithm independently on the two sets of student data. Four clusters were derived from the Fall 2019 cohort and three clusters were found in the Fall 2020 cohort. Descriptions of the clusters are found in Tables IV and V. The clusters for Fall 2019 confirm the previous finding that most students use their computers more frequently and in longer durations when accessing the LMS. However, cluster #2 describes a cohort of students (approximately 13.3% of the sample) that uses their mobile device to interact with the LMS more frequently than their computer. Further, cluster #2 has the highest rate of concurrent sessions — those where a student is interacting with the LMS on both a mobile device and a computer simultaneously.

The Fall 2020 clusters support the previous findings that students tended to have more frequent interactions with the LMS on both types of devices and that computer-based sessions were for longer durations. Cluster #0 in the Fall 2020 group

TABLE IV
CLUSTERS OF THE FALL 2019 STUDENT COHORT & MEAN (SD) FEATURE VALUES (N=24,354)

ClusterID: Description	*# of Students (%)*	*# Computer*	*# Mobile*	*# Concurrent*	*Duration - Computer*	*Duration - Mobile*
0: Computer sessions are 6x more than mobile	12,414 (51.0%)	162.4 (116.6)	25.9 (26.2)	3.8 (5.9)	4.6 (2.3)	1.0 (1.5)
1: Computer sessions are twice mobile; some concurrent sessions	8,321 (34.2%)	145.4 (100.3)	79.7 (58.1)	14.3 (15.0)	5.0 (2.3)	1.3 (1.1)
2: More mobile than computer sessions; significant concurrent sessions	3,238 (13.3%)	84.6 (76.0)	142.4 (130.0)	34.0 (43.9)	5.0 (3.0)	1.7 (1.2)
3: Very little LMS usage	381 (1.6%)	7.8 (7.3)	1.7 (3.0)	0.2 (0.6)	9.8 (7.5)	0.8 (3.8)

TABLE V
CLUSTERS OF THE FALL 2020 STUDENT COHORT & MEAN (SD) FEATURE VALUES (N=16,003)

ClusterID: Description	*# of Students (%)*	*# Computer*	*# Mobile*	*# Concurrent*	*Duration - Computer*	*Duration - Mobile*
0: More frequent mobile sessions than computer; significant concurrent sessions	3,073 (19.2%)	138.8 (127.4)	222.1 (203.0)	57.0 (63.0)	6.6 (3.5)	1.4 (1.1)
1: Few mobile sessions and few concurrent sessions	7,279 (45.5%)	232.7 (178.0)	36.6 (40.7)	7.5 (10.7)	5.8 (2.4)	0.8 (1.9)
2: Used both computer and mobile with some concurrent sessions	5,651 (35.3%)	210.3 (150.4)	122.6 (94.6)	26.3 (26.6)	6.2 (2.5)	0.9 (0.8)

TABLE VI
DERIVED ASSOCIATION RULES FOR FALL 2019

Cluster	Rule	C	S	L
0	Graduate Informatics, 25-29 ->Male, Asian	1.00	0.0005	10.26
0	Science, Male ->20-21	1.00	0.0005	2.47
1	Public Affairs, Female ->20-21	1.00	0.0014	2.91
1	Hispanic/Latino, Public Affairs, 18-19 ->Female	1.00	0.0007	2.05
2	Arch. & Design, Female, 18-19 ->White	0.94	0.0015	1.70
2	Black, Male, No Major ->18-19	0.94	0.0158	1.71
2	Hispanic/Latino, Female, No Major ->18-19	0.94	0.0138	1.70
3	Hispanic/Latino, No Major ->18-19	1.00	0.0210	3.23

TABLE VII
DERIVED ASSOCIATION RULES FOR FALL 2020

Cluster	Rule	C	S	L
0	Science, Black, Female ->20-21	1.00	0.0016	2.02
0	Hispanic/Latino, Public Affairs, Male ->20-21	1.00	0.0020	2.02
0	Public Health, White, 22-24 ->Female	1.00	0.0020	1.81
1	Hispanic/Latino, Female, Informatics ->20-21	1.00	0.0007	1.88
1	Engineering, White, 20-21 ->Male	1.00	0.0008	1.72
1	Technology ->Male	1.00	0.0007	1.72
2	22-24, No Major ->Male	1.00	0.0009	2.06
2	Arch. & Design, Asian ->Female, 18-19	1.00	0.0011	2.48

consists of 19.2% of the cohort who interacts with the LMS via a mobile device more often than they do on a computer. This cluster also shows the highest number of concurrent sessions and the longest mean duration of sessions on a mobile device. Interestingly, this cluster also has the longest mean duration of sessions on a computer although the frequency of those sessions is fewer than the other clusters. Cluster #2 (35.3% of the sample) also uses mobile devices frequently, although for shorter durations.

C. Association Rules Learned from the Clusters

Each set of clusters was examined to discover demographic patterns in the form of association rules. The specific demographics used when generating the rules include age group, school of major, ethnicity, and gender. All four demographics were found to contribute to the rules found within of the clusters in both the Fall 2019 and Fall 2020 groups. The rules derived from the Fall 2019 clusters show differences in the LMS behavior between older and younger students, for example. Students in the 18-19 age group were found more reliably in Fall 2019's clusters #1-3. Clusters #1 & #2 in Fall 2019 demonstrated higher usage of the LMS on mobile devices than cluster #1, so we might generalize that younger students tend to use their mobile devices more frequently. Students without a declared major (which often correlates with younger students) were also found in the clusters where mobile device use was higher.

The rules found in the Fall 2020 clusters saw more division between the school in which the student majors and their ethnicity, indicating that the student behaviors may have become more striated in the semester during the pandemic. Fall 2020's cluster #0 consists of some minority students —- including those in older age groups — as well as females, which indicates those groups may be more likely to use their mobile devices to access the LMS, or at least do so more frequently. Cluster #1 from Fall 2020 has the lowest usage of mobile devices and rules derived from this cluster demonstrate that the population trends toward male students who major in STEM subjects.

D. Revisiting Descriptive Statistics Based on Discovered Association Rules

The association rules discovered after the clustering provide insight to the specific demographic categories where behavioral differences might be found. For example, several Fall 2019 rules revealed patterns related to students with no declared major. Those rules indicated that students without a major were frequently found in Fall 2019's clusters #2 (higher mobile use than computer) & #3 (very little LMS use). Due to this discovery, we then conducted a statistical comparison of students without a declared major and students who have *any* major listed. A Mann-Whitney test indicated that students without a declared major had a higher median number of mobile-based sessions in Fall 2019 (Median=44) than students with any major declared (Median=35), U=56810711.0, $p<.001$ with a large effect size. However, another test found that was reversed in Fall 2020: students without a major declared (Median=46) had *fewer* sessions on a mobile device than students with any major declared (Median=62), U=5377451.5, $p<.001$ and a moderate effect size.

Further, the association rules allow us to discover combinations of demographic categories that might merit statistical comparisons. For example, Fall 2019's rules show various combinations of female students who are Black/African-American or Hispanic/Latino in the clusters where mobile device use is frequent. A Mann-Whitney test indicated that students who are female and either Black/African-American or Hispanic/Latino use their mobile device to interact with the LMS more (Median=41) than other students (Median=37), U=20187023.5, $p<.001$ and a large effect size. This difference is more pronounced in the Fall 2020 semester, where students who are female and either Black/African-American or Hispanic/Latino (Median=78) used mobile devices much more frequently than other students (Median=60), U=9670412.0, $p<.001$, with a large effect size.

V. Discussion

The findings presented in this study show that the transition to predominantly online courses in Fall 2020 affected the behavior of students' engagement with an LMS in terms of the devices used for access as well as the frequency and duration of those sessions. The average number of LMS sessions in high-enrollment STEM courses increased nearly 50% between the Fall 2019 and Fall 2020 semesters. This increase is logical, as students were primarily remote and accessing a wider variety of course materials online in the LMS. Of interest is the increase in the number of mobile sessions and events that were conducted via mobile devices during the Fall 2020 semester. Although students were largely absent from the campuses involved in this study and likely conducting their studies from home, they still appeared to rely on mobile devices to conduct academic work. One third of LMS sessions and nearly one quarter of all events in the LMS during the Fall 2020 semester were done so on a mobile device, and both of these numbers are statistically significant increases from the Fall 2019 semester. However, the duration of mobile sessions during Fall 2020 was shorter than in Fall 2019. This could indicate that students were reviewing the courses from a mobile device more frequently, but were doing so for short-duration tasks, such as responding to messages, reading requirements, or reviewing grades. These figures correspond with an observed increase in the use of mobile devices for academic work, but are particularly interesting given that the students were, in fact, *less* mobile during the Fall 2020 semester. More study is needed to determine if the increased reliance on mobile devices stems from necessity or simply choice.

Cluster analysis shows that there is no clear pattern of how students balance the use of computers and mobile devices. The cluster with the largest numbers of students (51.0%) in Fall 2019 shows that students overwhelmingly used computers more frequently than mobile devices. However, another cluster representing a significant 13.3% of students shows more frequent use of mobile devices than a computer to interact with the LMS. The clusters for Fall 2020 support the notion that students used their mobile devices more frequently — if for shorter durations — during the semester where 98.3% of their high-enrollment STEM courses were in an online format.

The findings also show that the behaviors of students and their personal technology does not fall into neat demographic patterns. However, association rule mining uncovered some demographics and demographic combinations that may be of particular interest for further research. Students without declared majors, who tend to be younger students, used their mobile devices more frequently during the Fall 2019 semester than they did when instruction moved online in Fall 2020. This change may be due to the fact that the younger students with no declared major had access to computers at home in

2020, or simply that there were fewer students who started their college career at the high point of the pandemic.

Female students, particularly those in underrepresented ethnicities, showed a tendency to use their mobile device for academic work more than other other students in Fall 2019 and that distinction grew more pronounced in Fall 2020. This could indicate that females either choose to use their mobile devices more frequently or, as supported by other studies, are more reliant on their mobile phones for academic work [31]. And, as discovered in student surveys at the outset of the transition to online learning, Black and Hispanic students were more likely to report issues with technology access and connectivity [32]. Thus, those students may have been more dependent on their mobile devices for academic work.

A. Limitations of this Study

This study analyzes only data from the LMS employed at the study site. While the LMS is the central point of connection between students and *most* of their learning technology, it is not the only source. Instructors at this study site employ a wide variety of institutional software (including videoconferencing, video hosting, and e-textbook platforms) that students can access directly without going through the LMS. Several of the courses included in this study also rely on content provided by textbook publishers which, again, can be accessed directly by students via bookmarks and apps. This means that this information can only inform the patterns of student behavior, rather than provide a comprehensive overview of it.

Further, this data cannot distinguish between sessions conducted on a laptop (which is portable) and one conducted on a desktop computer (which generally has a fixed location). Although laptops are functionally equivalent to desktop computers, they have the affordance of portability. There is no way to determine if the increase in mobile device use in Fall 2020 is aligned with a reduction in portability of devices in the computer classification used in this study.

VI. Conclusion

This study describe the collection, processing, and analysis of raw log data from an LMS integrated with student demographic data. The purpose of this study was twofold: to describe the data analysis process, and to determine if students' behavior changed with regard to their personal technology when comparing a semester prior to and a semester during the COVID-19 pandemic when most courses were offered online. We examined student behavior related to the device used to access the LMS, the events conducted while connected to the system, and the frequency and duration of access.

The study shows that students used the LMS much more frequently in the semester where their courses were primarily offered online. Students also used their mobile devices more frequently, although for shorter durations, during the semester with online courses. Demographic differences emerged, which warrants further study into how technological disparities between groups may influence their ability to successfully complete their academic coursework.

References

[1] G. Marinoni, H. v. Land, and T. Jensen, "The impact of COVID-19 on higher education around the world," International Association of Universities, Paris, France, Tech. Rep., Aug. 2020. [Online]. Available: https://bit.ly/3AZunQd.

[2] D. C. Gierdowski, "Student Experiences with Connectivity and Technology in the Pandemic," EDUCAUSE, Boulder, CO, Tech. Rep., Apr. 2021. [Online]. Available: https://www.educause.edu/ecar/research-publications/2021/student-experiences-with-connectivity-and-technology-in-the-pandemic/introduction-and-key-findings.

[3] D. C. Brooks, "Student Experiences Learning with Technology in the Pandemic," Tech. Rep., Apr. 2021. [Online]. Available: https://www.educause.edu/ecar/research-publications/2021/student-experiences-learning-with-technology-in-the-pandemic/introduction-and-key-findings.

[4] S. Grajek, *EDUCAUSE COVID-19 QuickPoll results: Help for students*, Apr. 2020. [Online]. Available: https://er.educause.edu/blogs/2020/4/educause-covid-19-quickpoll-results-help-for-students (accessed 03/25/2021).

[5] A. Aristovnik, D. Keržič, D. Ravšelj, N. Tomaževič, and L. Umek, "Impacts of the COVID-19 pandemic on life of higher education students: A global perspective," *Sustainability*, vol. 12, no. 20, p. 8438, 2020. DOI: 10.3390/su12208438.

[6] E. M. Aucejo, J. French, M. P. U. Araya, and B. Zafar, "The impact of COVID-19 on student experiences and expectations: Evidence from a survey," *Journal of Public Economics*, vol. 191, pp. 104 271–104 271, 2020, ISSN: 0047-2727. DOI: 10.1016/j.jpubeco.2020.104271.

[7] D. C. Gierdowski, "ECAR study of community college students and information technology, 2019," ECAR, Louisville, CO, Tech. Rep., May 2019. [Online]. Available: https://www.educause.edu/ecar/research-publications/ecar-study-of-community-college-students-and-information-technology/2019/introduction-and-key-findings.

[8] S. Sarkar, "The role of information and communication technology (ICT) in higher education for the 21st century," *The Science Probe*, vol. 1, no. 1, pp. 30–40, May 2012, ISSN: 2777-9566.

[9] R. Castro, "Blended learning in higher education: Trends and capabilities," *Education and Information Technologies*, vol. 24, no. 4, pp. 2523–2546, 2019, ISSN: 1360-2357. DOI: 10.1007/s10639-019-09886-3.

[10] V. Mathew, A. I. Abduroof, and J. Gopu, "Transforming higher education through digitalization," in *Transforming Higher Education Through Digitalization : Insights, Tools, and Techniques*, S. L. Gupta, N. Kishor, N. Mishra, S. Mathur, and U. Gupta, Eds., Boca

Raton, FL: CRC Press, Oct. 2021, pp. 145–171, ISBN: 9781000461299. DOI: 10.1201/9781003132097-9.

[11] B. Kang, "How the COVID-19 pandemic is reshaping the education service," in *The future of service post-COVID-19 pandemic, volume 1*, ser. The ICT and Evolution of Work, J. Lee and S. H. Han, Eds., Feb. 2021, ISBN: 9789813341258. DOI: 10.1007/978-981-33-4126-5.

[12] R. Seilhamer, B. Chen, S. Bauer, A. Salter, and L. Bennett, *Changing mobile learning practices: A multiyear study 2012–2016*, en, 2018. [Online]. Available: https://er.educause.edu/articles/2018/4/changing-mobile-learning-practices-a-multiyear-study-2012-2016 (accessed 11/03/2020).

[13] J. Mella-Norambuena, R. Cobo-Rendon, K. Lobos, F. Sáez-Delgado, and A. Maldonado-Trapp, "Smartphone use among undergraduate STEM students during COVID-19: An opportunity for higher education?" *Education Sciences*, vol. 11, no. 8, p. 417, 2021. DOI: 10.3390/educsci11080417.

[14] J. Pomerantz, M. Brown, and D. C. Brooks, "Foundations for a next generation digital learning environment: Faculty, students, and the LMS," en, ECAR, Louisville, CO, Tech. Rep., 2018, p. 16. [Online]. Available: https://library.educause.edu/-/media/files/library/2018/1/ers1801.pdf.

[15] C. B. Mpungose and S. B. Khoza, "Postgraduate students' experiences on the use of Moodle and Canvas learning management system," *Technology, Knowledge and Learning*, pp. 1–16, 2020, ISSN: 2211-1662. DOI: 10.1007/s10758-020-09475-1.

[16] J. Schoonenboom, "Using an adapted, task-level technology acceptance model to explain why instructors in higher education intend to use some learning management system tools more than others," *Computers & Education*, vol. 71, pp. 247–256, 2014, ISSN: 0360-1315. DOI: 10.1016/j.compedu.2013.09.016.

[17] D. Laurillard, "The pedagogical challenges to collaborative technologies," *International Journal of Computer-Supported Collaborative Learning*, vol. 4, no. 1, pp. 5–20, 2009, ISSN: 1556-1607. DOI: 10.1007/s11412-008-9056-2.

[18] *Canvas LMS REST API documentation*. [Online]. Available: https://canvas.instructure.com/doc/api/ (accessed 04/02/2022).

[19] K. R. Suneetha and R. Krishnamoorthi, "Identifying user behavior by analyzing web server access log file," *IJCSNS International Journal of Computer Science and Network Security*, vol. 9, no. 4, pp. 327–332, Apr. 2009.

[20] D. Clow, "An overview of learning analytics," *Teaching in Higher Education*, vol. 18, no. 6, pp. 1–13, 2013, ISSN: 1356-2517. DOI: 10.1080/13562517.2013.827653.

[21] A. Rubel and K. M. L. Jones, "Student privacy in learning analytics: An information ethics perspective," *The Information Society*, vol. 32, no. 2, pp. 143–159, 2016, ISSN: 0197-2243. DOI: 10.1080/01972243.2016.1130502.

[22] C. Hodges, S. Moore, B. Lockee, T. Trust, and A. Bond, *The difference between emergency remote teaching and online learning*, Mar. 2020. [Online]. Available: https://er.educause.edu/articles/2020/3/the-difference-between-emergency-remote-teaching-and-online-learning (accessed 04/14/2022).

[23] *Classification of Instructional Programs (CIP)*. [Online]. Available: https://nces.ed.gov/ipeds/cipcode/default.aspx?y=56 (accessed 04/02/2022).

[24] *Eligible CIP Codes for the STEM OPT Extension*. [Online]. Available: https://studyinthestates.dhs.gov/stem-opt-hub/additional-resources/eligible-cip-codes-for-the-stem-opt-extension (accessed 04/02/2022).

[25] *User-Agent - HTTP*. [Online]. Available: https://developer.mozilla.org/en-US/docs/Web/HTTP/Headers/User-Agent (accessed 04/02/2022).

[26] *Solved: API & Mobile Device Usage*. [Online]. Available: https://community.canvaslms.com/t5/Canvas-Mobile-Users/API-Mobile-Device-Usage/td-p/154895 (accessed 04/02/2022).

[27] R. Pegoraro, "What Is a Chromebook?" *U.S. News & World Report*, Mar. 2021. [Online]. Available: https://www.usnews.com/360-reviews/technology/laptops/what-is-a-chromebook (accessed 03/03/2022).

[28] M. Banerjee, "An exploratory study of online equity: Differential levels of technological access and technological efficacy among underserved and underrepresented student populations in higher education," *Interdisciplinary Journal of E-Skills and Lifelong Learning*, vol. 16, pp. 93–121, Feb. 2020. DOI: 10.28945/4664.

[29] P. J. Rousseeuw, "Silhouettes: A graphical aid to the interpretation and validation of cluster analysis," *Journal of computational and applied mathematics*, vol. 20, pp. 53–65, 1987.

[30] R. Agrawal, T. Imieliński, and A. Swami, "Mining association rules between sets of items in large databases," in *Acm sigmod record*, ACM, vol. 22, 1993, pp. 207–216.

[31] A. L. Gonzales, J. M. Calarco, and T. Lynch, "Technology problems and student achievement gaps: A validation and extension of the technology maintenance construct," *Communication Research*, vol. 47, no. 5, pp. 750–770, 2020, ISSN: 0093-6502. DOI: 10.1177/0093650218796366.

[32] S. S. Jaggars, B. A. Motz, M. D. Rivera, *et al.*, "The digital divide among college students: Lessons learned from the COVID-19 emergency transition," Midwestern Higher Education Compact, Tech. Rep., Mar. 2021, p. 20. [Online]. Available: http://files.eric.ed.gov/fulltext/ED611284.pdf.