THE PENNSYLVANIA STATE UNIVERSITY
SCHREYER HONORS COLLEGE

COLLEGE OF INFORMATION SCIENCES AND TECHNOLOGY

LEARNING ANALYTICS: A CASE STUDY OF A WEB-ENHANCED COURSE

BRITTNEY ROMANOWSKI
SPRING 2016

A thesis
submitted in partial fulfillment
of the requirements
for a baccalaureate degree
in Information Sciences and Technology
with honors in Information Sciences and Technology

Reviewed and approved* by the following:

Abdullah Konak
Professor of Information Sciences and Technology
Thesis Supervisor

Jeanne Rose
Associate Professor of English
Honors Adviser

* Signatures are on file in the Schreyer Honors College.

ABSTRACT

Online learning has grown steadily over the past few years, and the use of learning analytics has grown in parallel. As online education continues to grow, so does the need to improve student learning online and the need to understand students' interaction in their electronic learning environment. This thesis first explored previously conducted research on online learning environments and the learning analytics tools implemented on them, as well as online environments that Google Analytics has been implemented in. This study then explored and tested the implementation of Google Analytics as a learning analytics tool on a course website at Penn State Berks, IST 210. Google Analytics was used to record student event data on the course website in order to better understand how the students interacted with it. In order to analyze patterns and trends, the data recorded by Google Analytics was exported into a statistical analysis tool, SPSS. Each of the pages on the IST 210 website was categorized by five main interactive characteristics. This categorization framework allowed for certain student usage behaviors to be compared to the different types of pages. This thesis shows how the application of Google Analytics can be a valid tool used to observe and improve student learning online.

TABLE OF CONTENTS

LIST OF FIGURES

LIST OF TABLES

ACKNOWLEDGEMENTS

First and foremost, I would like to thank my mom, Kimberly Romanowski. Mom, you have been an outstanding role model. You taught me to push myself to my limits and then exceed them. Thank you for teaching me that it's okay to be encountered with a difficult challenge; that is the only way we are going to grow and develop as a person. This thesis was by far my greatest educational challenge. Thanks again mom. I love you.

Next, I would like to thank Dr. Feinstein, my original honors advisor, for introducing me to the Schreyer Honors College program. Thank you for letting me know that you had faith in me and thought that if anyone was capable of becoming a Schreyer Scholar, it was me.

Dr. Konak, I would like to thank you for being an outstanding thesis advisor and working with me for hours on end. You taught me what the true essence of research was, a skill that I will cherish throughout my entire professional career.

Last but not least, I would like to thank Dr. Rose, my current honors advisor. Thank you for jumping on board my senior year and picking right up where Dr. Feinstein left off; you did it so seamlessly. Thank you so much for keeping me on track my senior year and making sure I met all deadlines. Thank you for reassuring me the entire way and keeping me cool, calm, and collected.

Chapter 1

Introduction

Online Learning

With increased development of information technology and the Internet, online learning has been revolutionized. As a result of this revolution, online learning has become a key component to the educational system. Findings by Allen and Seaman (2014) state that the total number of students taking at least one online course has reached a new high of 7.1 million students, accounting for 33.5% of the total enrollment of degree-granting post-secondary institutions (Allen & Seaman, 2014). According to a survey conducted by the Babson Survey Group in January 2014, the methods educators are using for online learning are Web Facilitated Courses with 1 to 29% of content delivered online; Blended/Hybrid Courses with 30 to 79% of content delivered online; and Online Courses with 80 to 100% of content delivered online (Allen & Seaman, 2014). The survey also comments on the fact that 90% of academic leaders believe that it is “likely” or “very likely” that in roughly five years a majority of higher education students will be enrolled in at least one online course (Allen & Seaman, 2014).

Until recently, researchers and instructors have relied on observing students in the classroom in a face-to-face setting in order to understand how they interact with their learning environment and course material. However, instructors today are also frequently using online content in traditional courses to enhance students’ experiences as well as broaden their access to information. In particular, some instructors are using online textbooks or online homework for students to use on their own at home, such as the course in this study. In addition, other instructors are using the pedagogical method of a flipped classroom, where students watch and review lectures before class time, so that instructors can focus on exercises during the class time.

With the emergence of online courses and course materials as well as the increased use of online content in traditional courses, instructors do not always have the liberty to observe their students in a face-to-face setting. Therefore, there is a pressing need for monitoring students' progress and capturing their interaction with online learning content. Learning Analytics is a promising solution to address this need.

Learning Analytics Background

Analytics is about understanding, interpreting, and communicating data. In broad terms, Baker and Siemens (2014) define analytics and data mining as "methodologies that extract useful and actionable information from large datasets." According to the National Agenda for Action Analytics, analytics involves "processes of data assessment and analysis that enable us to measure, improve, and compare the performance of individuals, programs, departments, institutions or enterprises, groups or organizations, and/or entire industries" (Norris, Baer, & Offerman, 2009). Data analytics has been used by a vast array of disciplines for decades. For instance in the Psychology field, the *Journal of Analytical Psychology*, first published in 1955 and still in publication today, is one of the earliest journals devoted primarily to analytics; similarly, the *Computers in Biology and Medicine Journal* has been publishing papers primarily focusing on analytics in biological sciences since 1970 (Baker & Siemens, 2014).

Conversely, in the education discipline there has been a lag in the emergence of research in analytics. It was not until 2009 that the *Journal of Educational Data Mining,* the first educational journal devoted primarily to analytics, was published (Baker & Siemens, 2014). This new variation of analytics is known as learning analytics (LA). LA was chosen to serve as the backbone for this study because of its growing attention to and importance in the field of education.

LA refers to the collection and analysis of relevant data that students create when they interact with their learning environment. The First International Conference on Learning Analytics and Knowledge in 2011 defined learning analytics as “the measurement, collection, analysis and reporting of data about learners and their contexts, for purposes of understanding and optimizing learning and the environments in which it occurs.” This definition directly relates to a definition by Siemens and Long (2011) where they describe learning analytics as a field that “centers on the learning process (which includes analyzing the relationship between learner, content, institution, and educator)” (p. 34). With the increase in distance learning and online courses as well as with the increased online content use in traditional courses, learning analytics offers a new methodology that can in turn change the way instructors design and teach their courses. Moreover, the increase in awareness and recognition of online learning methods are resulting in an increase in the traceability of LA.

Learning Analytics Tools

Currently, there are a number of competing analytic tools that specialize in analyzing the educational data created by students. Some commercially available Course Management Systems, such as BlackBoard, Ellucian, Canvas, and Angel, have the ability to record student data, incorporating the objectives behind LA tools. These systems are also recognized as Learning Management Systems (LMS), Course Management Systems (CMS), and Student Systems. The type of data that these CMS are collecting in large is mainly student event data.

The analytics tool configured on the website collects and records student usage data that can later be analyzed for trends, patterns, irregularities, or indicators that would generate attention to the data. These findings are then interpreted by analysts or instructors educated in the area of analytics, and suggestions for improvements to the site or course material can be made. The process of collecting and interpreting the student usage data can lead to improved student learning through the efficiency and simplification of the mediums used.

Aim of Study

The main problem that instructors face today is an absence of new methods and tools that can record and observe student usage data. These new methods and tools will need to be cohesive with current pedagogical methodology in order to fully optimize an instructor's observation methods for online courses. The primary focus of this study is to determine whether Google Analytics, a more general analytics tool, can be customized to aid in the evaluation of the learning system or course website based on the information that Google Analytics gathers. Can Google Analytics be used to identify the areas of the learning system or course website that need to be improved? The secondary focus of this study is to determine whether Google Analytics can collect data that can be used to establish a connection between student behavior and their performance.

To analyze the potential applicability of Google Analytics, this study surveys previous scholarship on learning analytics and reviews the current tools and methods associated with them. In the background investigation, current methods and tools available for learning analytics are identified and contrasted. Previous case studies on learning analytics are then reviewed to assess the types of data that were collected and analyzed and to examine how that data were used. Following the literature review, this study provides a case study on the implementation of Google Analytics into one course website at the Penn State Berks campus, IST 210.

Chapter 2

Learning Analytics

Literature Review

The research on learning analytics revealed a distinct trend. A significant amount of research focused on investigating the correlation of student-recorded events (such as participation in both required and optional activities, frequency of sessions, and length of sessions) with the overall success rate of students in the course. Romero-Zaldivar, Pardo, Burgos, and Delgado Kloos (2012) studied the relationship between the learning activities that students performed in a virtual computer workspace and their achievement in an engineering course. They reported that six (Command, Worktime, Compile, Editor, Profiler, and IDE) of the eight observed student event types were found to have positive and significant correlations with the student's final grade.

In a separate study by Valenzuela & Whale (2012), a similar connection was found. The results of 123 students who completed one subject were analyzed; of the 123 students 117 passed and 6 failed. They found that 91% of the students who completed the activity observed (which was scheduled early in the teaching period) were more likely to successfully complete the subject. However, out of the remainder of the students that did not complete the activity only 72% successfully completed the subject.

Additionally, to further support this trend is a pilot program at Purdue University, titled Course Signals. In this pilot program, Arnold and Pistilli (2012) studied the relationship between the implementation of Signals and the students' overall success in the course. Signals is a system that was built specifically to run as a learning analytics tool. Signals' main purpose is to calculate a student's risk factor based on four categories, which are performance, effort, prior academic history, and student characteristics (Arnold & Pistilli, 2012). After the pilot program was implemented, students saw varying success. The courses in the pilot program showed an increase in performance. Individual courses varied from 2.23% receiving A's and B's to other course showing an increase by 13.84% of students receiving A's and B's (Arnold & Pistilli, 2012). Additionally, courses also saw a decrease in C's. Some courses saw a 1.84% decrease in C's, while others saw a 9.38% decrease. Finally, there was a similar trend with a decrease in students receiving D's and F's. One course saw a .59% decrease, while others saw as high as a 9.40% decrease (Arnold & Pistilli, 2012).

Additionally, supporting this trend were case studies that focused their data analysis on the correlation of student-recorded events with predicting student success and retention. Particularly, research by Bainbridge, Melitski, and Zahradnik (2015) supported this trend. Bainbridge et al. (2015) recorded student event data and attributes and predicted students at risk. Attributes measured were the students' gender, cumulative GPA, and the number of times that students performed certain activities such as reading and posting forums, reading the course content, accessing the course content, and logging in (Bainbridge et al., 2015). They used the quantitative data collected by the learning analytics tool to generate the student's academic risk, and their model had an 84.84% accuracy rate (Bainbridge et al., 2015). Further, Campbell, DeBlois, and Oblinger (2007) summarized other cases where institutions have utilized data collection to predict student success.

Romero-Zaldivar et al. (2012) also focused on predicting student success and retention using learning analytics. They performed a multiple regression analysis to predict future student success based on student-recorded events. After establishing the correlations between the student-recorded events, the authors then selected two events, the "WorkTime" and "Profiler" to build their linear regression model. Their model was able to account for more than 20% of the variability of student final grades (Romero-Zaldivar et al., 2012). To conclude, the authors of the study discovered that the data obtained using the monitoring procedure could be used to develop a linear model that can predict the final course grade (Romero-Zaldivar et al., 2012).

Chapter 3

Google Analytics

Background

Google Analytics is a web analytics tool that offers Data Collection & Management, Data Consolidation, Data Analytics & Reporting, and Data Activation ("Our history in depth," 2013). The purpose of this application is to analyze, track, and measure website traffic. Google Analytics is a versatile tool that offers unique metrics (also known as quantitative measurements) to monitor how the user is interacting with the website. Some examples of general data that can be collected from Google Analytics are user data, session data, traffic sources, platform or device used to access site, page tracking, content grouping, site speed, social interactions, app tracking, event tracking, and many more (Dimensions & Metrics Explorer). Google Analytics can show what types of devices users are accessing the site with, such as desktop computer, mobile device, or tablet. From an instructional design perspective, this information could give great insight into how to design the layout of the website based on how students are accessing it. Once the metrics are gathered, Google Analytics then organizes the information into a user-friendly, easy to understand way. It accomplishes this with charts, graphs, and other tools.

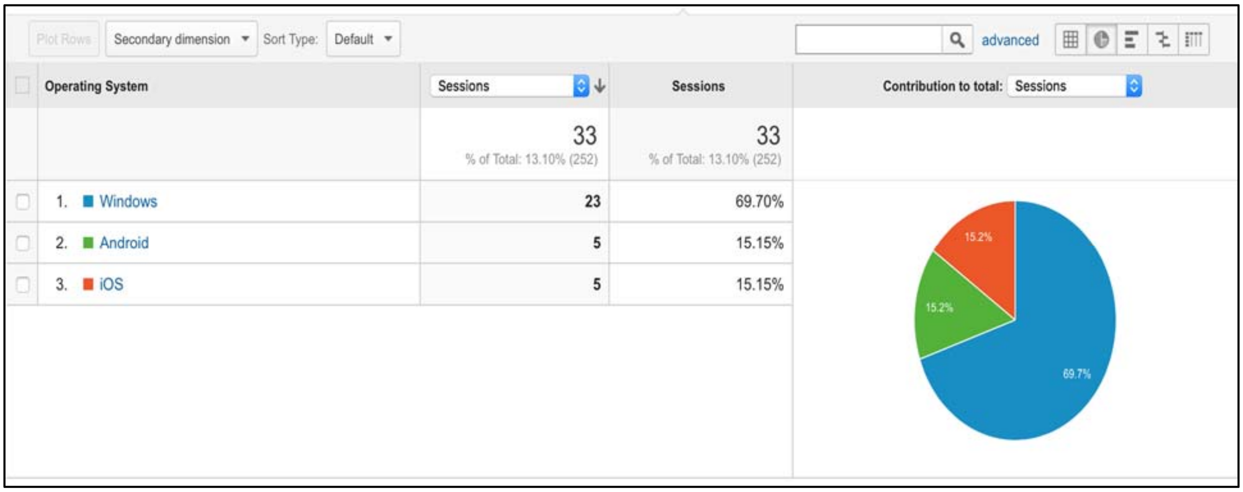

Figure 3-1 Google Analytics Operating System Report

Figure 3-1 illustrates the breakdown of the operating systems that users access the site from, displayed as a pie chart. The primary dimension used in Figure 3-1 is the "operating system" in the mobile category. The metric "session" in Figure 3-1 is the number of users per operating systems per the total sessions.

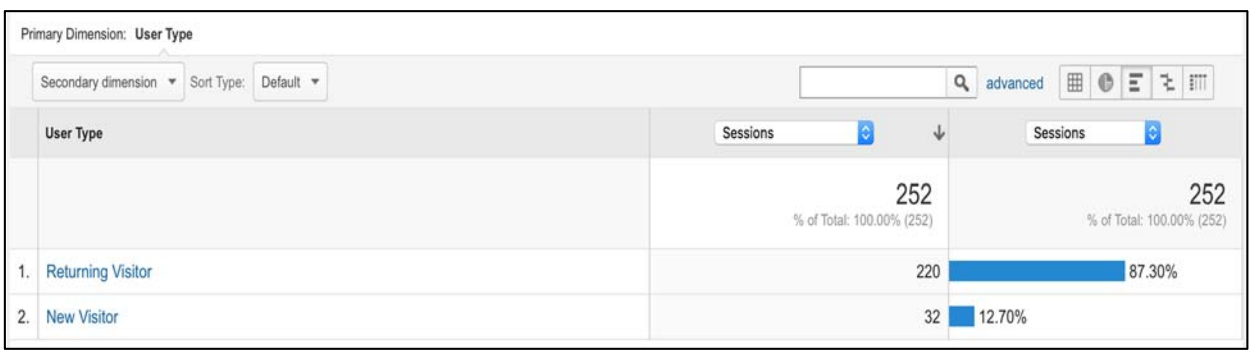

Figure 3-2 Google Analytic User Type Report

Figure 3-2 displays session data from November 12 through December 12, 2015. Of the 252 initiated sessions, 220 were returning users and 32 were new. The primary dimension used in the figure is "user type" (returning or new users) in the behavior category and the metric used was the number of "sessions" executed by each user type.

Data Types

Google Analytics separates its data into two types: dimensions and metrics. Dimensions describe the characteristics of the users, their sessions, and actions. Metrics are similar to dimensions except they describe the quantitative measurements of users, sessions, and actions. Every report contains both types of data. Google Analytics takes on a much more dynamic approach to the type of data recorded compared to the commercially available LA tools mentioned previously. For instance, Google Analytics will automatically track how students are interacting with the site through session engagement.

Dimensions	**Description**
Pageviews	Pageviews is the total number of pages viewed. Repeated views of a single page are counted.
Unique Pageviews	Unique pageviews is the number of sessions during which the specified page was viewed at least once. A unique pageview is counted for each *page URL + page Title* combination.
Average Time on Page	The average amount of time users spent viewing a specified page or screen, or set of pages or screens.
Entrances	New versus returning users, frequency & recency, engagement, user-ID coverage, site speed, site search, site content
Bounce Rate	Bounce rate is the percentage of single-page visits (i.e., visits in which the person left your site from the entrance page without interacting with the page).
% Exit	% exit is (number of exits) / (number of pageviews) for the page or set of pages. It indicates how often users exit from that page or set of pages when they view the page(s).
Browser	The browsers used by visitors to your website.
Operating System	The operating systems used by visitors to your website. Includes mobile operating systems such as Android.

Table 3-1 Google Analytics Dimensions Description

Table 3-1 summarizes the dimension types Google Analytics offers as well as some of the corresponding metrics available to us for our case study.

For example, Google Analytics will track numerous interaction elements, such as what pages the student viewed, components they clicked on those pages, the path they took throughout their session, as well as bounce rates (“Features”). The term “bouncing” refers to a user landing on a page and then leaving the page immediately with no interaction with it. From a learning perspective, Carini, Kuh, and Klein (2006), argue that it is important for students to be engaged with their course material as it has been shown to have a positive correlation to their learning. Therefore, if the web page has a high bounce rate, then students may not be engaging enough with it. Many of the features mentioned within Google Analytics are not available with the analytics tools built into course management systems. For example, ANGEL CMS takes a more static approach to collecting student interaction data, such as the session length, if they completed assigned tasks, and logon dates, but not actually how the students are interacting with the course pages.

Literature Review

In a study conducted on the Rutgers University Newark Library, Feng (2007) implemented Google Analytics to track the library's users' behavior and pinpoint the users' motivation behind the information that they sought. The library had used a number of different methods to track visitors' behavior over the past five years but found their results with Google Analytics to be especially promising. Feng (2007) found Google Analytics' visual features to be of the utmost help in analyzing their data. The Trend Reporting Feature of Google Analytics allowed researchers to compare data from different date ranges. For example, they were able to compare trends from before and after the website redesign (Feng, 2007). Another tool they utilized was the Defined Funnel Navigation; they were able to determine how many users accurately followed paths they had designed to reach target pages (Feng, 2007). After the redesign was completed, pre- and post-redesign site data were compared, and Feng (2007) found that the library site had been improved in the following ways: the site now brings in more traffic, the site achieves better user loyalty (with more returning users), and finally the site has improved user navigation.

A separate study by Arendt and Wagner (2010) found similar success with implementing Google Analytics to redesign the web page of Morris Library at Southern Illinois University Carbondale. In their case study, they distinguished four types of dimensions available through Google Analytics they used in aiding their website redesign. Those dimensions are basic reports, most popular content, navigation summary, and keywords (Arendt & Wagner, 2010). Findings from the basic reports metric of Google Analytics informed the library that 95% of their users had a screen resolution of 1024 X 768, meaning that they did not need to design the pages to also include lower resolutions (something that had been done in the library's last redesign in 2005) (Arendt & Wagner, 2010). Through the use of Google Analytics, user navigation confusion was discovered. The library found that users were generally getting confused on the functions of the "eJournal Finder" section and the "Databases / Find Articles" section of the website. The confusion was discovered based on the percentage of expected navigation paths of users and actual navigation paths of users. The library clarified cues to make the purposes of those sections better understood (Arendt & Wagner, 2010).

Particularly, Clark, Nicholas, and Jamali (2014) studied Google Analytics' role and ability to track user behavior and website effectiveness. Clark et al. (2014) argue that Google Analytics has already been successfully used by researchers and would be highly recommended by those researchers to study the users' behavior, effectiveness of the website, and web traffic. In order to test the effectiveness of Google Analytics' ability to evaluate a website's performance, Clark et al. (2014) referenced two studies. In one study, Plaza (2011) tested Google Analytics' effectiveness in analyzing data from a website about the urban regeneration of the city of Bilbao during a time span from February 4, 2007 until January 30, 2010. During the specified time period, Plaza (2011) found that Google Analytics was a sufficient tool in evaluating website performance.

Chapter 4

Methodology

Course Website Background

The website being evaluated by Google Analytics is the course website for IST 210: Organization of Data, at Penn State University's Berks Campus. The website is intended to be an online interactive textbook, which aims to bring databases to life with a unique, problem-based approach. The website is used as the sole textbook resource for IST 210 students, which allows them to have access to their textbook anywhere they have access to the Internet. The course website is hosted through a website called WordPress. The course is taught in a computer lab, so every student will have access to a computer during class time. For students who are using a tablet or other mobile device the course website also has a mobile friendly design. The website has in-page exercises and assessments built into the section pages, allowing students to practice and test their knowledge as they go. Each page of the website has Google Analytics' tracking code built into it, allowing the analytic tool to capture in-page statistics through the use of cookies and JavaScript. The tracking code is linked to an account created with Google Analytics that allows the account owners to view the statistical data.

How WHERE works

Let us think about how the WHERE clause actually works. The comparison statement given in the WHERE clause is evaluated for each row of the table, and then the query returns only the rows ***where*** the comparison statement is true. For example, in the query

```
SELECT *
FROM SHIPMENT
WHERE Quantity>= 200;
```

Quantity>=200 is TRUE in rows (2, 3, 8, 9, 10) as shown in Figure 1.; hence, only these rows are returned when WHERE Quantity >=200 is used with the SELECT statement.

SNo	Vendor	Part	City	SDate	Quantity	Method	UnitCost	Quantity >=200
1	V1	P1	Reading	2006-09-20	100	Online	10.5	0
2	V1	P2	Reading	2006-10-27	300	Online	20	1
3	V1	P3	Reading	2006-11-01	200	Phone	10.5	1
4	V2	P1	York	2006-11-23	100	Phone	5	0
5	V2	P3	York	2007-01-01	120	Phone	3.5	0
6	V3	P4	Lancaster	2007-01-10	120	Store	23	0
7	V3	P1	Lancaster	2007-01-25	130	Store	30	0
8	V2	P2	York	2007-02-05	210	Online	40.5	1
9	V1	P4	Reading	2007-02-05	230	Online	12.3	1
10	V1	P4	Lancaster	2007-02-05	230	Online	30	1

Figure 1. Evaluation of the Quantity >=20 comparison statement

Practice Questions:

- List the shipping no, vendor, part, and city of online shipments and sort the output in the descending order of shipping date.
- List the shipping no and city of the shipments which are not for York.

Figure 4-1 Course Website Practice Questions Example

Figure 4-1 is a snapshot from the IST 210 website demonstrating the in-page exercises available at the end of a section. These practice questions allow the students to put the material into perspective with hands on activities.

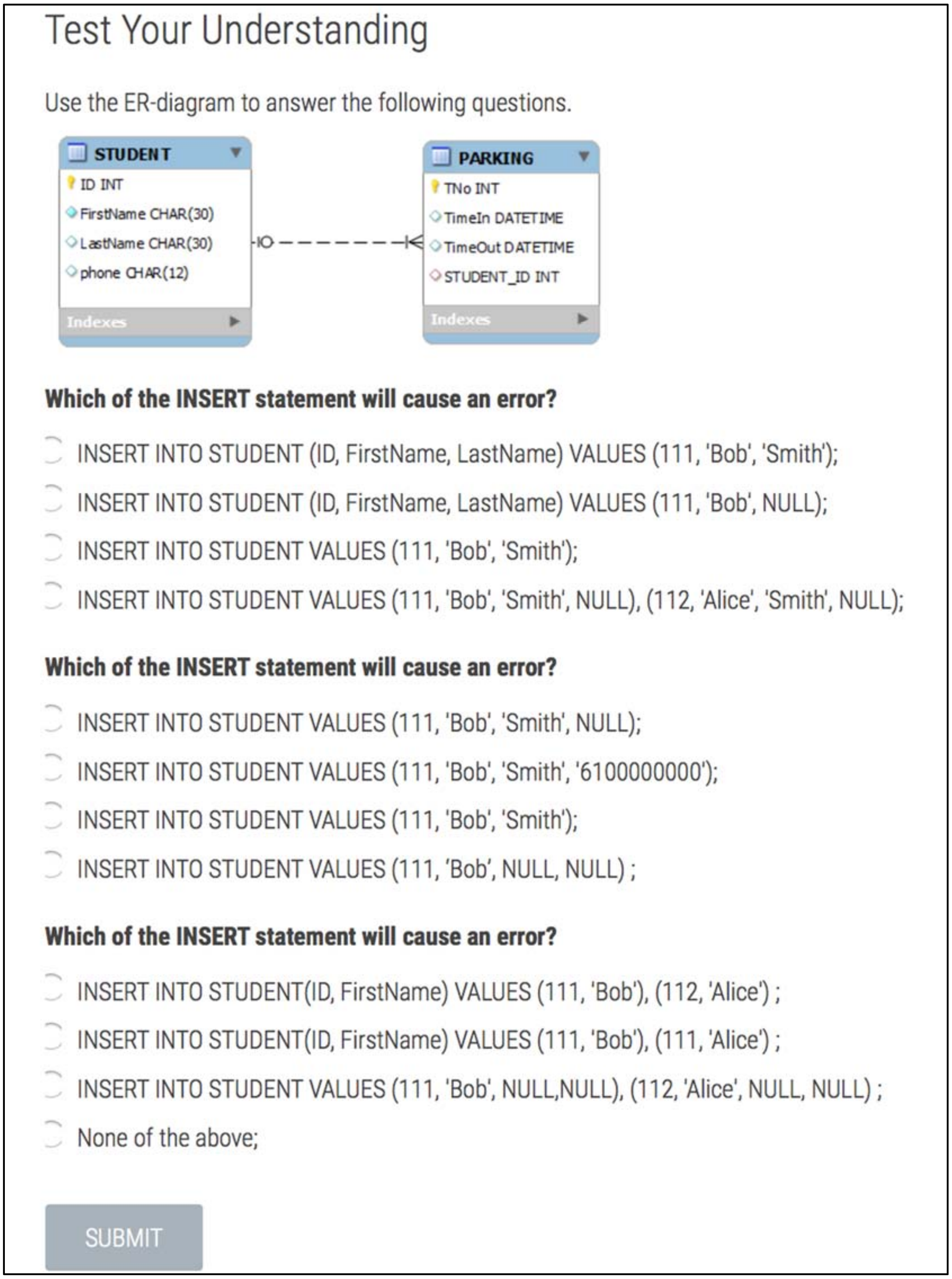

Figure 4-2 Course Website Quiz Example

Figure 4-2 is an example of an in-page assessment that a student would find after reading and practicing the section material. The quizzes are not graded and can be taken an unlimited number of times. The in-page assessments are intended to help students by allowing them to test their understanding of the material and receive instant feedback on their performance.

Participants

Penn State's University Bulletin describes IST 210 Organization of Data as a course that "aims to prepare students for obtaining fundamental understanding on the database concepts and practical skills to analyze and implement a well-defined database design." It accomplishes this goal by providing, "an introduction to physical database design, data modeling, relational model, logical database design, SQL query language, and instructors' choices on database applications and advanced concepts."

IST 210 has a prerequisite of IST 110: Information, People, and Technology. IST 210 is an entrance to major requirement for students majoring or minoring in Information Sciences and Technology. An entrance to major course requires students to receive a C or better. The course meets in person twice a week. The first section meets from 12:15 PM to 1:30 PM, and the second section meets from 4:30 PM to 5:45 PM on Tuesdays and Thursdays, and, as mentioned previously, the textbook is the course website. As an introductory IST course, IST 210 is generally comprised of first-year and second-year students. The participants in this study are 43 of the students in both sections of the course. Of the 44 total enrolled students, 6 females and 37 males agreed to participate. While this course has a traditional face-to-face meeting, it relies heavily on a web aspect for the book and student involvement.

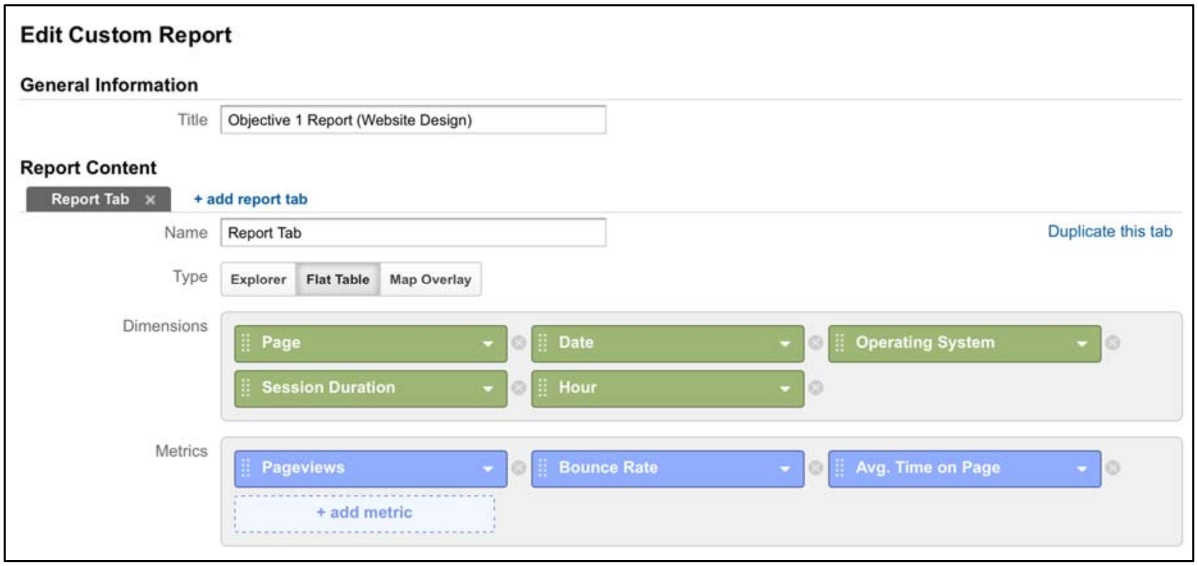

Figure 4-3 Google Analytics Edit Custom Report

Figure 4-3 is the administrator view when creating a custom report in Google Analytics. This figure shows the dimensions and metrics that were chosen for this study.

	Page	Date	Operating System	Session Duration	Hour	Pageviews	Bounce Rate	Avg. Time on Page
1.	/database/5-data-access/5-2-conditional-statements/	20160301	Macintosh	3988 seconds	03	8 (0.56%)	0.00%	00:01:06
2.	/database/5-data-access/5-1-select-order-by/	20160301	Macintosh	3988 seconds	03	7 (0.49%)	0.00%	00:02:48
3.	/database/4-normalization/2nf/	20160218	Windows	4127 seconds	12	6 (0.42%)	0.00%	00:07:44
4.	/database/5-data-access/5-2-conditional-statements/	20160229	Windows	2759 seconds	21	6 (0.42%)	0.00%	00:03:34
5.	/database/4-normalization/4-3-3nf/	20160218	Windows	2396 seconds	12	5 (0.35%)	0.00%	00:06:35
6.	/database/5-data-access/5-2-conditional-statements/	20160228	Windows	1702 seconds	13	5 (0.35%)	0.00%	00:03:45
7.	/database/5-data-access/5-2-conditional-statements/	20160301	Windows	1228 seconds	10	5 (0.35%)	0.00%	00:03:01
8.	/database/	20160223	Windows	0 seconds	12	4 (0.28%)	100.00%	00:00:00
9.	/database/2-conceptual-design/2-1-db-design-process/	20160229	Windows	580 seconds	21	4 (0.28%)	0.00%	00:00:25
10.	/database/4-normalization/4-4-2nf-and-3nf-examples/	20160223	Macintosh	473 seconds	03	4 (0.28%)	0.00%	00:01:28

Figure 4-4 Google Analytics Custom Report Output

Figure 4-4 shows the outputted result of running the custom report in Google Analytics.

Data Collection

After each weekly meeting of the IST 210 sections, the data recorded by Google Analytics were downloaded and exported into an Excel file. Additionally, the data were generated into a custom report in order to analyze dimensions and metrics that are relevant to the study's objective of determining whether Google Analytics can be customized to aid in the evaluation of the learning system or course website.

The individual dimensions chosen were page, date, operating system, session duration, and hour. The individual metrics that were chosen for the report were pageviews, bounce rates, and average time on page, as previously discussed in Chapter 3. The chosen dimensions and metrics for this study were particularly important to analyze for the primary objective, which was to determine whether Google Analytics, a more general analytics tool, could be customized to aid in the evaluation of the learning system or course website based on the information that Google Analytics gathers. The main metrics used for analysis were pageviews and average time on page. These metrics were used to compare each web page and the content on the page to the projected amount of student interaction. A total of 2297 data points were generated between the time period of January 10, 2016 through March 4, 2016. The analysis only included the pages that contained course materials covered during the specified time period, which were Chapters 1 through 5. Additionally, the first day of class was excluded from the analysis as students were instructed to go on the course website to become familiar with it and ask questions. This day was excluded because students were not looking at the website for content; rather, they were becoming familiar with its layout, generating inaccurate student behavior data for the context of this study.

Additional Statistical Analysis Tool

For the purpose of this study, the weekly data downloaded from Google Analytics was imported into the statistical software SPSS. To test our first objective, *t*-tests were executed to understand the relationship between the users' interactions with the pages by comparing bounce rates, average session times, and average page time. A *t*-test allows the user to view any statistically significant difference in the means of two groups of data. These findings are reported in Chapter 5.

Page Coding

In order to analyze the students' relationship with the IST 210 course website, each of the website's pages needed to be categorized based on level of interaction. The purpose of categorizing the website's pages was to determine the relationships between the data collected (dimensions) and the interaction attributes of the website's pages through *t*-test and, as described in Chapter 5, ANOVA test. Thereby, the attributes of pages that engaged students the most could be determined.

In order to categorize the pages, each page was carefully observed for a number of characteristics. The pages were coded based on five attribute categories: quizzes, exercises, code examples, images, or videos. If a page contained the content, it would receive a (1) in that category, and (0) otherwise. After the page was analyzed for all categories, its scores were added. Each page could receive a total of five points if it contained all the categories. The lowest score that the page could receive was 0 points if the page only had text and contained none of the five categories.

In addition to the above categorization, the pages were also examined further for the number of words, images, and videos that each page contained as well. To count the words on each page, the page settings were accessed in each page's edit view. The website administrator was able to view a word count for each page, which indicated the length of the page.

Chapter 5

Results

The results of this study on Google Analytics followed the same trend as the case studies reviewed earlier on other Learning Analytics tools. The primary objective in this study focused on determining whether Google Analytics, a more general analytics tool, can be customized to aid in the evaluation of the learning system or course website based on the information that Google Analytics gathers. Through the analysis of the data collected, Google Analytics provided sufficient student event data to understand students' relationship with the course website. The first step in understanding the students' behavior and relationship with the course website was to understand when students are accessing it.

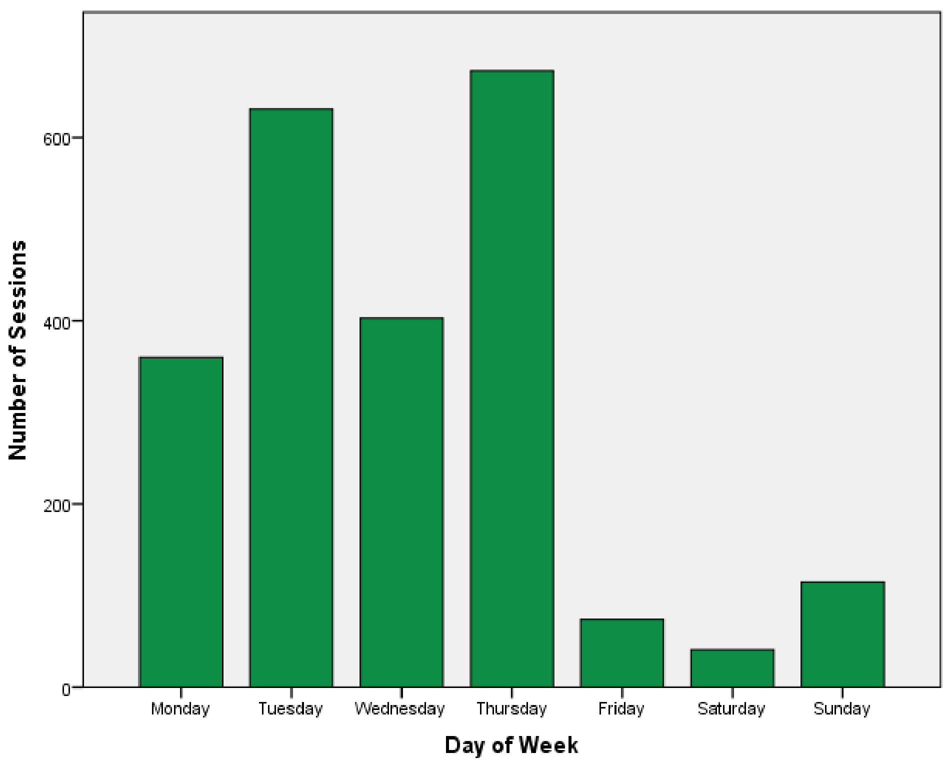

Figure 5-1 Day of Week Students Access Course Website

Figure 5-1 Illustrates each day of the week and the total number of sessions initiated on those given days.

Figure 5-1 indicates that on Tuesdays and Thursdays, the days of the course meetings, the course website was the most active, having the greatest sessions initiated on those days. Figure 5-1 also shows that Mondays and Wednesdays, the days directly before class meets, have the second highest amount of sessions for the week. This correlation shows us that students are often on the course website the day before class, interacting with course material. Next, in order to further understand when the students were accessing the course website, the hour of the day most sessions were being initiated was examined.

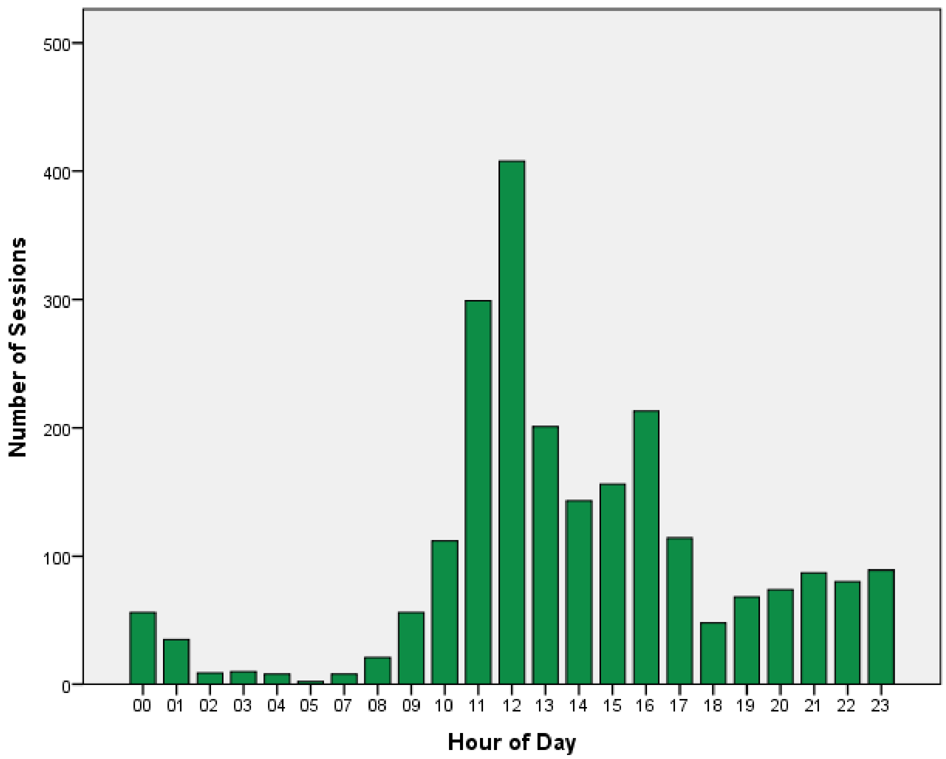

Figure 5-2 Hour of Day Students Access Course Website

Figure 5-2 Illustrates each hour of the day and the total number of sessions initiated on those given hours.

Figure 5-2 shows that students are accessing the course website the most during the class meeting times (12:15 PM and 4:30 PM), which is expected. However, the figure also shows that students are visiting the course's website primarily right before class (when assignments are due) and later in the evening (increasing after 6:00 PM until midnight). Assignments are typically due before the beginning of class, which is at 12:15 PM for the first section. The graph shows us that students are accessing the website in increasing numbers the hour before and leading up to the submission deadline. For the second section starting at 4:30 PM, the graph shows the same trend. Therefore, it can be concluded from the graph that students primarily access the course website during class hours and directly before class starts when assignments are due. This trend was also verified by the CANVAS course management data indicating that the majority of the students tend to submit their work within the last hour of the due time.

What types of pages students are viewing the most and what types of pages they spent most of their time on was the next piece to understand in order to get a better picture of the students' behavior. The types of pages that students are viewing the most was an important characteristic to look at because it may provide insight on the type of learning material students gravitate towards and are stimulated by.

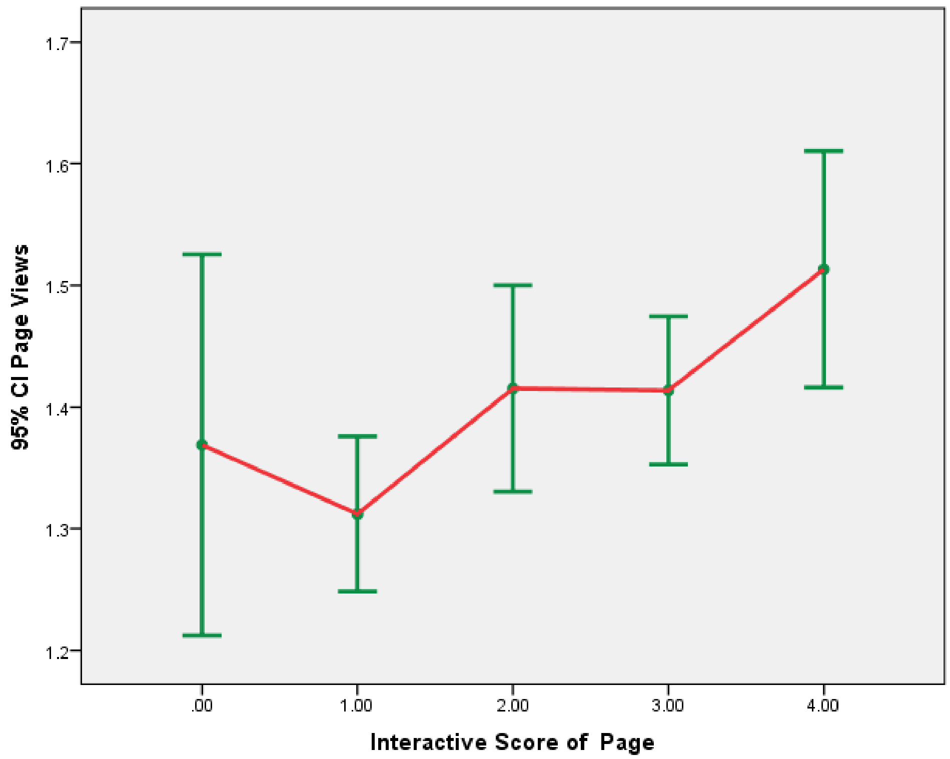

Figure 5-3 Pageviews and Interactive Score Graph

Figure 5-3 illustrates the relationship between the pageviews and the interactive score of a page.

The data collected revealed a significant correlation between the number of views a page received and the interactive score of the page. Based on the page's interactive score it was determined that the greater the interactive score the more views the page received, as shown in Figure 5-3. Pages that have an interactive score of .00 are solely text-based, but still contain important course content. However, it can be concluded that students gravitate more towards website pages that are less text-based and require more student engagement. However, students spend a significant amount of time on pages that are solely text-based. The content within the solely text-based pages focuses more on terms and concepts that make up the course content.

Students can be graded on this aspect and are therefore likely to prioritize it. Approximately 20% of all exams and quizzes come directly from pages that are solely text-based.

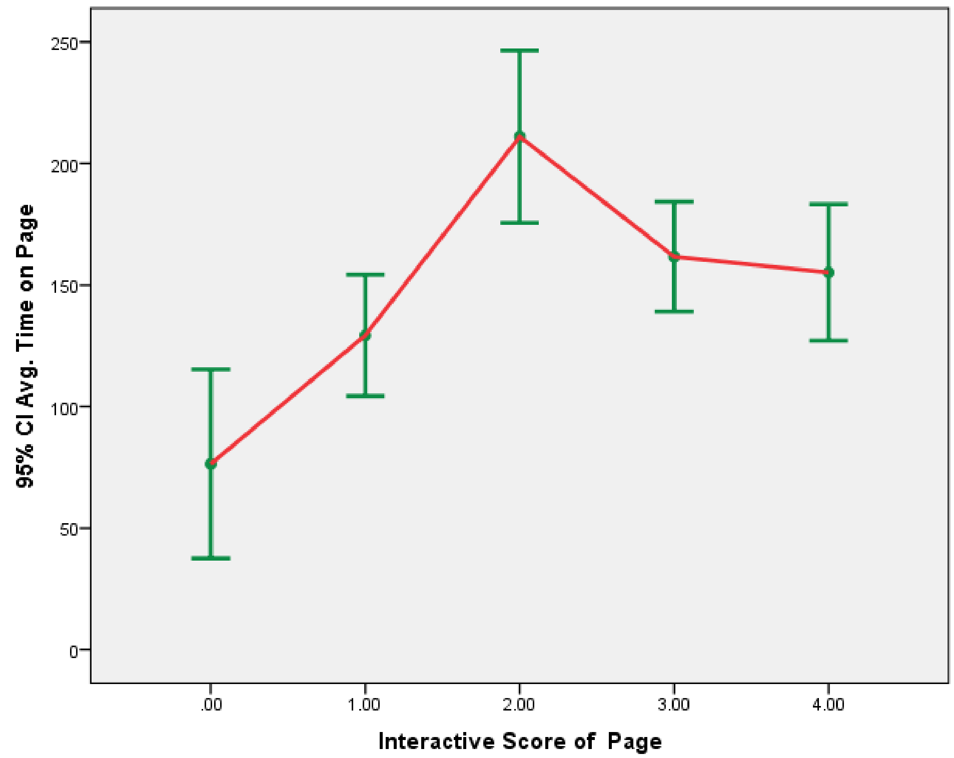

Figure 5-4 Average Time on Page and Interactive Score Graph

Figure 5-4 illustrates the relationship between the average time on page (seconds) and the interactive score of page.

Figure 5-3 shows that the higher the interactive score of a page the more views that page received. However, Figure 5-4 shows that the higher the interactive score of a page does not always indicate the greater time spent on that page. Figure 5-4 shows that from an interactive page score of 0 to 2 there is a positive correlation between average time spent on a page and that page's interactive score. After an interactive score of 2 this finding has a negative correlation. One reason for this relationship could be that pages that have an interactive score are mainly pages with quizzes. The next analysis conducted was on how many words per page compared to the average time the students would spend on those pages.

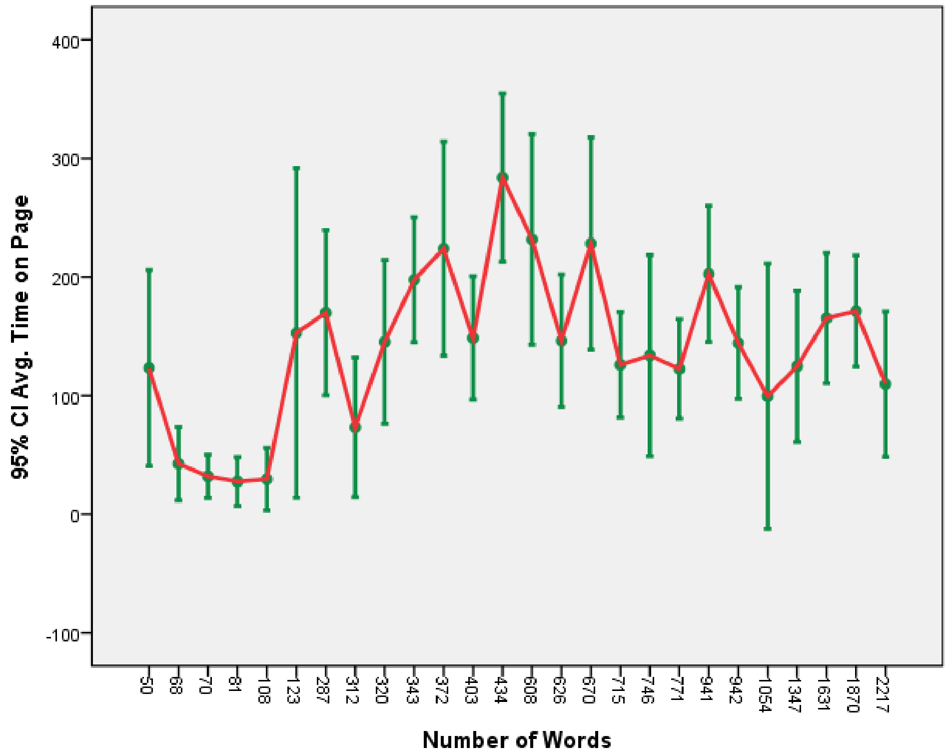

Figure 5-5 Average Time on Page and Number of Words per Page

Figure 5-4 illustrates the relationship between the number of words on the course website's individual web pages and the average time spent on pages with the corresponding amount of words.

The results from this analysis were particularly important to this study. The analysis revealed having roughly 435 words per page seems to reflect an optimal page length level for students. The following tables illustrate numerically the statistical results of comparing the students' interactions with pages that have or do not have the four individual interactive components that make up a page's interactive score.

Characteristics/Involvement		**Mean**	**Std. Deviation**	***t* value**	***p* value**
Average Time on Page	Without Quiz	136.16 Seconds	298.849	-3.238	0.001
	With Quiz	179.99 Seconds	343.137		
Pageviews	Without Quiz	1.36 Views	0.876	-2.411	0.016
	With Quiz	1.45 Views	0.897		

Table 5-1 The Effect of Quizzes on Student Engagement

Table 5-1 shows the means and standard deviations of the metrics average time on page and pageviews for the pages with and without a quiz. For average time on page, there is a strong significant and positive relationship between the presence of a quiz and average time spent on that page (t=-3.238, p=0.001). For pageviews, there is also a significant and positive relationship between the presence of a quiz and pageviews (t=-2.411, p=0.016). These results show that having an interactive quiz in the page has a positive effect on students' engagement with the page.

Characteristics/Involvement		Mean	Std. Deviation	*t* value	*p* value
Average Time on Page	Without Exercises	157.68 Seconds	318.479	0.610	0.542
	With Exercises	162.91 Seconds	332.783		
Pageviews	Without Exercises	1.42 Views	0.882	-1.704	0.089
	With Exercises	1.39 Views	0.900		

Table 5-2 The Effect of Exercises on Student Engagement

Table 5-2 does not show any significant relationship between the average time on page and the presence of an exercise on that page (t=0.610, p=0.542). In addition, the relationship between pageviews and the presence of an exercise on that page was not significant (t=-1.704, p=0.089).

Characteristics/Involvement		Mean	Std. Deviation	*t* value	*p* value
Average Time on Page	Without Images	129.43 Seconds	291.503	-2.377	0.018
	With Images	168.24 Seconds	332.247		
Pageviews	Without Images	1.39 Views	0.925	-.399	0.690
	With Images	1.41 Views	0.878		

Table 5-3 The Effect of Images on Student Engagement

In Table 5-3, the relationship between the presence of images and student engagement are analyzed. There is a significant relationship between the average time on page and the presence of images (t=-2.377, p=0.018). For pageviews, however, no significant relationship was observed (t=-0.399, p=0.690).

Characteristics/Involvement		Mean	Std. Deviation	*t* value	*p* value
Average Time on Page	Without Videos	163.53 Seconds	334.962	0.855	0.393
	With Videos	150.96 Seconds	297.601		
Pageviews	Without Videos	1.37 Views	0.825	-3.353	0.001
	With Videos	1.50 Views	1.014		

Table 5-4 The Effect of Videos on Student Engagement

In Table 5-4, the relationship between the presence of videos and student engagement are analyzed. There is a significant relationship between the pageviews and the presence of videos (t=-3.353, p=0.001). For average time on page, however, no significant relationship was observed (t=0.855, p=0.393).

Characteristics/Involvement		Mean	Std. Deviation	*t* value	*p* value
Average Time on Page	Without Code examples	163.01 Seconds	333.991	0.610	0.542
	With Code examples	154.55 Seconds	307.947		
Pageviews	Without Code examples	1.38 Views	0.828	-1.704	0.089
	With Code examples	1.45 Views	0.976		

Table 5-5 The Effect of Code Examples on Student Engagement

Table 5-5 does not show any significant relationship between the average time on page and the presence of a code example on that page (t=0.610, p=0.542). In addition, the relationship between pageviews and the presence of a code example on that page was not significant (t=-1.704, p=0.089).

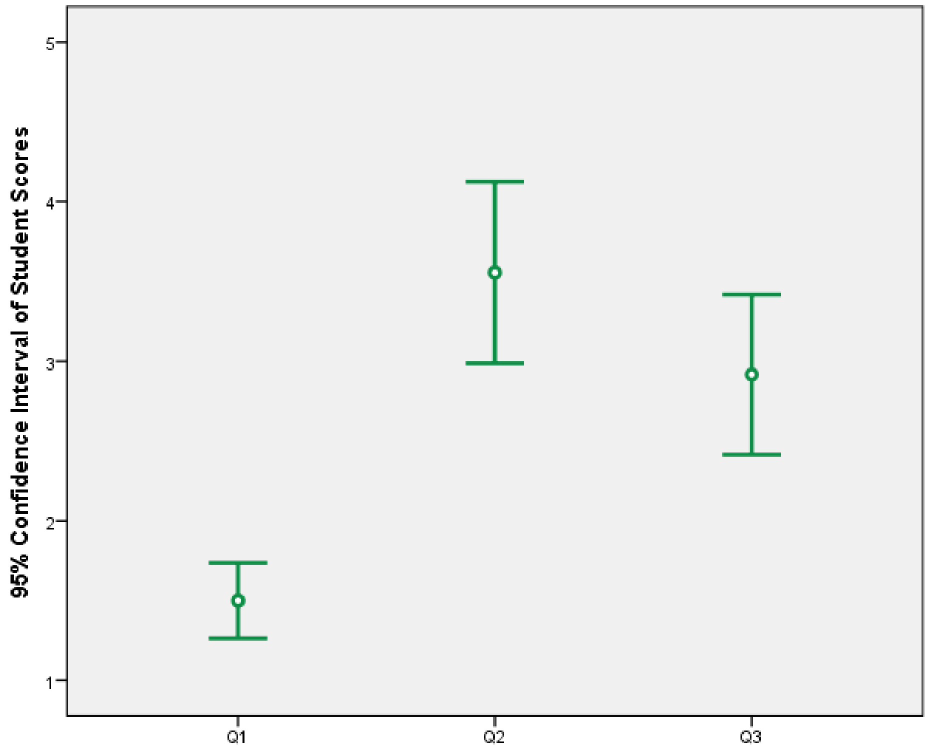

Figure 5-6 Survey Results of Student Study Strategies

Figure 5-6 shows how students use the website to get ready for the test on a Likert-scale from 1-strongly agree to 5-strongly disagree. Q1 represents the survey question of, “When I review the course notes (web or pdf), I try some of the quizzes to make sure that I understand them.” Q2 represents the survey question of, “When I review the course notes (web page or pdf), I practice design and normalization examples without looking at the notes.” Q3 represents the survey question of, “When I review the course notes (web page or pdf), I try to write query examples without looking at the notes.” This figure shows that (for student engagement) quizzes are the most important feature. Students do not necessarily practice the examples in the website’s pages, but they tend to take the quizzes and try to learn their responses.

The tables above show the statistical difference and significance between the average time on page and pageviews for the course's web pages that have one of the interactive components incorporated or not. These components are the ones used to determine a page's interactive score. The tables that have a *p* value less than 0.05 shows significance in the relationship of the compared components. In addition to analyzing the students' average time on page with the pageviews through a *t*-test for each of the independent factors, an Analysis of Variance (ANOVA) was conducted on all of the components together, using the total interactive score of the pages.

		Sum of Squares	**Df**	**Mean Square**	**F**	**Sig.**
Average Time on Page	Between Groups	3583280.419	10	358328.042	3.448	0.073
	Within Groups	237566678.976	2286	103922.432		
	Total	241149959.395	2296			
Pageviews	Between Groups	13.456	10	1.346	1.709	0.000
	Within Groups	1799.686	2286	0.787		
	Total	1813.142	2296			

Table 5-6 ANOVA

That means being analyzed through the ANOVA test above are those gathered from Tables 5-1 through 5-4, the main components that make up interaction scores. Table 5-5 shows a statistically significant difference between the means of the pageviews' groups, with a *p* value of 0.000. However, table 5-5 does not show a statistically significant difference between the means of the average time on page's groups with a *p* value of 0.073. Nonetheless, this result is very close to being significant. It should be noted that average time on a page had a very large variability.

Chapter 6

Discussion and Implications

This study was intended to test a general analytics tool and determine whether or not it could perform the basic functions of a learning analytics tool. The need for this study stemmed from a lack of current options for instructors to observe their students' behavior outside of the traditional classroom setting. The current methods did not allow for an instructor to dynamically observe how their students interacted with their online content. In addition, there were not many options for instructors who have a course website outside of the institution's course management system.

Through the research this study has conducted, it has been determined that Google Analytics, in fact, was an effective alternative to a learning analytics tool. Google Analytics successful fulfilled this study's primary objective of determining whether Google Analytics, a more general analytics tool, can be customized to aid in the evaluation of the learning system or course website based on the information that Google Analytics gathers. Google Analytics was proven to gather enough student event data that could then in turn be analyzed to understand the students' behavior when interacting with the course website. This resulted in an understanding of how to better tailor the course website to the students' learning.

The information presented from this study's analysis resulted in some distinct findings about students' behavior with the course website. First, text density has a significant impact on the amount of time a student will stay interested in a page. Findings from this study showed that the optimal page length for students is about 435 words per page. From this finding, it was concluded that to best suit students' learning, it would be beneficial to break long pages into separate pages or sections, with more interactive content integrated. However, it is important to not overstimulate the student at the same time, but to maintain an equal balance of engaging activities to reading content.

An additional finding from this study was what interactive feature students find the most beneficial when going back and reviewing course material. Students expressed that the in-page quizzes were the main feature within the course website they use for reviewing. These quizzes offer instant feedback to the students in order for them to gauge their understanding of the material. The data collected through Google Analytics confirmed this claim because students tend to spend more time on a page if the page includes a quiz. Through student feedback, it is recommended that instructors that have course websites have in-page quizzes built in to the content. Students can take these in-page quizzes on their own, to test their understanding and review the material, which they found valuable. According to students, exercises were not as helpful as they did not provide instant feedback and did not allow them to self-assess their understanding.

The study was not able to determine if Google Analytics can collect data that can be used to establish a connection between student behavior and their performance, the secondary objective. The objective was unable to be determined because the course website is hosted through WordPress, which is made accessible to instructors through Penn State. When using an institutionalized version of WordPress, instructors do not have all the features of a non-institutionalized user. This version did not allow for code and other features to be changed to accommodate the objective of tracking individual users. A non-institutionalized user would have the ability to change code within WordPress and could implement an individual tracking ID for each user. This could then allow for the individual analysis of each user's behavior with the site.

However, this study shows that Google Analytics can be used to collect data for better designing course websites and identify webpage attributes that engage students the most. Such data combined with data gathered through course management systems can be used to enhance student learning in online and traditional courses that depend heavily on online content. Based on the continually growing percentages of students involved in online and hybrid courses, this study shows the capabilities of Google Analytics and the significant impact it can have in the field of education. Google Analytics offers instructors and professionals alike the opportunity to implement an analytics tool that allows them to monitor and understand the behavior the users have with their website, leading to potential optimization.

BIBLIOGRAPHY

Allen, E., & Seaman, J. (2014). Grade change: Tracking online education in the United States. Retrieved September 24, 2015, from http://onlinelearningconsortium.org/survey_report/2013-survey-online-learning-report/

Arendt, J., & Wagner, C. (2010). Beyond description: Converting web site usage statistics into concrete site improvement ideas. *Journal of Web Librarianship*, *4*(1), 37-54.

Arnold, K. E., & Pistilli, M. D. (2012, April). Course Signals at Purdue: Using learning analytics to increase student success. In *Proceedings of the 2nd International Conference on Learning Analytics and Knowledge* (pp. 267-270). ACM.

Bainbridge, J., Melitski, J., & Zahradnik, A. (2015). Using learning analytics to predict at-risk students in online graduate public affairs and administration education.

Baker, R., & Siemens, G. (2014) Educational data mining and learning analytics. In Sawyer, K. (Ed.) *Cambridge Handbook of the Learning Sciences: 2nd Edition*, pp. 253-274.

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

Carini, R. M., Kuh, G. D., & Klein, S. P. (2006). Student engagement and student learning: Testing the linkages. *Research in Higher Education*, *47*(1), 1-32.

Clark, D. J., Nicholas, D., & Jamali, H. R. (2014, July). Evaluating information seeking and use in the changing virtual world: The emerging role of Google Analytics [Electronic version]. *Learned Publishing*, *27*(3), 185-194. doi:10.1087/20140304

Dimensions & metrics explorer. (n.d.). In *Google Developers*. Retrieved January 5, 2016, from https://developers.google.com/analytics/devguides/reporting/core/dimsmets

Features. (n.d.). Retrieved October 16, 2015, from https://www.google.com/analytics/standard/features/

Feng, W. (2007, June 1). Using Google Analytics for improving library website content and design: A case study. *Library Philosophy and Practice (e-journal)*. Retrieved from http://digitalcommons.unl.edu/cgi/viewcontent.cgi?article=1121&context=libphilprac

IST 210 Organization of Data. (n.d.). In *University Bulletin*. Retrieved February 2, 2016, from http://bulletins.psu.edu/bulletins/bluebook/university_course_descriptions.cfm?letter=I&courselong=ist%7C210%7Clatest

Norris, D., Baer, L., & Offerman, M., (2009). "A national agenda for action analytics," White paper of outcomes from the National Symposium on Action Analytics, held September 21–23 in St. Paul, Minnesota (November 15, 2009), 1.

"Our history in depth." 2013. *Google Company*. Retrieved October 17, 2014 from, http://www.google.com/about/company/history/

Plaza, B. (2011). Google analytics for measuring website performance. *Tourism Management*, *32*(3), 477-481.

Romero-Zaldivar, V., Pardo, A., Burgos, D., & Delgado Kloos, C. (2012;2011). Monitoring student progress using virtual appliances: A case study. *Computers & Education, 58*(4), 1058-1067. doi:10.1016/j.compedu.2011.12.003

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

Valenzuela, F. R., & Whale, S. (2012). Enhancing distance education student outcomes utilizing learning analytics: A case study.

ACADEMIC VITA

Brittney Romanowski
Bwr5170@psu.edu

Education
B.S in Information Sciences and Technology with a focus in design and development. Minor in Security Risk Analysis with a focus in network security.
Honors in Information Sciences and Technology
Thesis Title: Learning Analytics: A Case Study of a Web-Enhanced Course
Thesis Supervisor: Abdullah Konak

Work Experience
June 1 – August 10, 2015
Solution Engineer Intern
Automated manual test scripts in the C# language through the use of coded UI and visual studio.
Deloitte Consulting LLC, 300 Corporate Center Dr., Camp Hill, PA 17011
Supervisor's Name: Mike Lapitan

January – May, 2015
RSVP IST Intern
Ran the project's website and implemented telecommunications for advisory board meetings to connect members from different states to meetings remotely.
Penn State University, 2080 Tulpehocken Rd, Reading, PA 19610
Supervisor's Name: Jayne Leh

January 2015 – May, 2016
Teaching Assistant
Assisted professor by maintaining course pages on the Angel Course Management System.
Created drop boxes for students to submit assignments, grade chapter notes, and keep attendance records.
Penn State University, 2080 Tulpehocken Rd, Reading, PA 19610
Supervisor's Name: William Bowers

Honors and Awards
Outstanding Internship in IST (2016)

IST Leadership Award (2016)

Boscov Honors Scholarship (2014-2016)

Dean's List (2012-2016)

Professional Memberships: Phi Kappa Phi

Community Service

Camp Noah (July 26 – August 2, 2014)

Head Soccer Coach for LYSC U-6 Soccer (2015 – Present)