Using Google Analytics to Improve the Course Website of a Database Course

Brittney Romanowski and Abdullah Konak
Deloitte Consulting LLP *Penn State Berks*

Abstract

Online learning has grown steadily in the last decade, and the use of learning analytics has increased in parallel. As online education continues to grow, instructors need to find new ways to enhance student learning online and to understand students' interactions with their electronic learning environment. This paper presents an implementation of Google Analytics as a learning analytics tool on a database course website. The course website was created as an interactive e-book, and the objective of the study was to discover which features of the website were most effective in improving student learning. During a semester, Google Analytics was used to record student event data on the course website in order to understand how students interact with the website. The collected data were analyzed to discover patterns and trends in student interactions. Discovered patterns were then correlated with various attributes of the individual website pages such as the level of interactivity and page content type. Findings showed that interactivity of a course page was the most important factor for increasing student engagement with the course content. In particular, in-page quizzes were found to be very effective in improving student engagement with the website. This preliminary study has shown how Google Analytics could be a valid tool to observe and improve student learning online.

Keywords

Learning Analytics, Google Analytics, Online Learning.

Introduction

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[1] 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. According to a survey[1] 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. 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.

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 as well.[5] In particular, some instructors are using online textbooks or online homework. 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.

Analytics is about understanding, interpreting, and communicating data. In broad terms, Baker and Siemens[2] define analytics and data mining as "methodologies that extract useful and actionable information from large datasets." Learning Analytics (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[3] 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."

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. The main objective of the paper is to investigate whether Google Analytics can be used to identify the areas of the learning system or course website that need to be improved.

Google Analytics

Google Analytics is a web analytics tool that offers Data Collection & Management, Data Consolidation, Data Analytics & Reporting, and Data Activation. 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.[4] Google Analytics can show what types of devices users are accessing the site with, such as desktop computers, mobile devices, or tablets. 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.

Google Analytics separates its data into two types: dimensions and metrics. Dimensions describe the characteristics of the users, their sessions, and actions as summarized in Table 1. Metrics are similar to dimensions except that 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 learning analytics tools. For instance, Google Analytics will automatically track how students are interacting with the site through session engagement.

Table 1- Google Analytics Dimensions

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.

Methodology

Target Website and Participants

The target website being evaluated by Google Analytics is the course website for a database course (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 course website was developed on the WordPress platform provided by the university. 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. 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.

As an introductory IST course, IST 210 is generally comprised of first-year and second-year students. The participants in this study are 44 of the students in both sections of the course. While this course has a traditional face-to-face meeting, it relies heavily on a web aspect for the textbook and student involvement.

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. 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 collected 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.

In order to analyze the students' relationship with the IST 210 course website, each of the website's pages were categorized based on the level of interactive elements that the page included. The purpose of categorizing the website's pages was to determine the relationships between the collected data (dimensions) and the interaction attributes of the website's pages. 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 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.

Summary of the Findings

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 1. 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. 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.

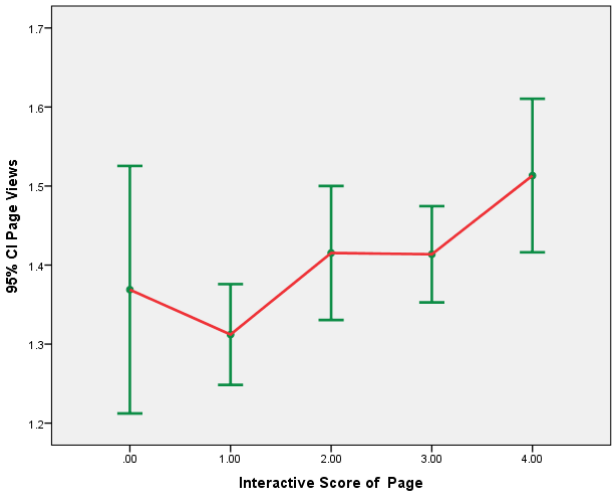

Figure 1. Pageviews and Interactive Score Graph

Figure 1 shows that the higher the interactive score of a page the more views that page received. However, Figure 2 shows that the higher the interactive score of a page does not always indicate the greater time spent on that page. One reason for this relationship could be that pages that have an interactive score of two are mainly pages with quizzes.

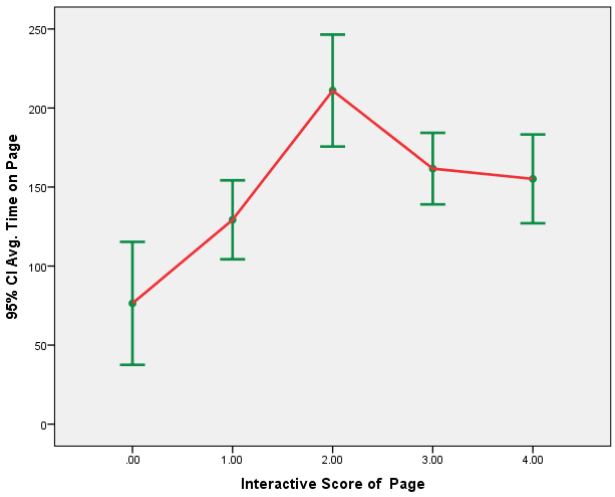

Figure 2. Average Time on Page and Interactive Score Graph

Table 2 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.

Table 2. The Effect of Quizzes on Student Engagement

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		

In addition to quizzes, some pages include exercises that students are expected to complete as they read through the text. Unlike quizzes, correct answers to exercises are not provided to students. Table 3 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).

Table 3. The Effect of Exercises on Student Engagement

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		

In Table 4, 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). It should be noted that illustrative images are frequently used in the course website to explain many database concepts. The results in Table 4 may show that students took time to study these illustrative images.

Table 4. The Effect of Images on Student Engagement

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		

In addition to quizzes, exercises, and images, two other interactive components, videos and code examples, did not have a statistically significant effect on Pageviews and Average Time on Page. For the brevity of the presentation, the statistics for these interactive items are not provided in this paper.

At the end of the semester, students also completed a survey to evaluate the course website. Figure 3 illustrates how students use the website to get ready for the tests 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 necessary practice the examples in the website, but they tend to take the quizzes and try to learn their responses.

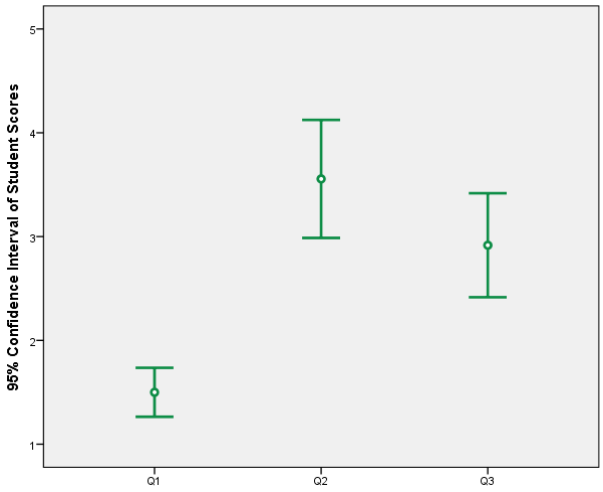

Figure 3. Survey Results of Student Study Strategies

Discussions and Implications of the Findings

In this preliminary study, it has been determined that Google Analytics is an effective alternative as a learning analytics tool. Google Analytics can 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. However, Google Analytics could not capture data pieces that can be directly associated with individual students. This is a drawback of the version of Google Analytics used in the study. 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.

The information presented from this study's analysis resulted in some distinct findings about students' behavior with the course website. The first finding from this study was what interactive feature students find the most beneficial when they review course material. Students expressed that the in-page quizzes were the main feature within the course website they used 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.

Conclusion

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.

References

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

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

3 The First International Conference on Learning Analytics and Knowledge, Retrieved September 24, 19, 2016, from https://tekri.athabascau.ca/analytics/

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

5 Konak, A., Ryoo, J., and Kulturel-Konak, S. "Student Perceptions of a Hands-on Delivery Model for Asynchronous Online Courses in Information Security," ***Proceedings of ASEE Mid-Atlantic Section Fall 2014 Conference***, Swarthmore College, Swarthmore, PA, November 14-15, 2014, 1-7.