

Cost Per User: Analyzing EZProxy Logs for Assessment

Tiffany M. LeMaistre
Nevada State College, tiffany.lemastre@nsc.edu

Follow this and additional works at: <https://docs.lib.purdue.edu/charleston>



Part of the [Collection Development and Management Commons](#), and the [Information Literacy Commons](#)

An indexed, print copy of the Proceedings is also available for purchase at:

<http://www.thepress.purdue.edu/series/charleston>.

You may also be interested in the new series, Charleston Insights in Library, Archival, and Information Sciences. Find out more at: <http://www.thepress.purdue.edu/series/charleston-insights-library-archival-and-information-sciences>.

Tiffany M. LeMaistre, "Cost Per User: Analyzing EZProxy Logs for Assessment" (2015). *Proceedings of the Charleston Library Conference*.
<http://dx.doi.org/10.5703/1288284316291>

Cost Per User: Analyzing EZProxy Logs for Assessment

Tiffany M LeMaistre, Electronic Resources and Discovery Librarian, Nevada State College

Abstract

Cost per use has long been a staple of collection development decision-making for electronic resources, but what of the users behind those retrieval and search counts? Questions about the interdisciplinary usage of an e-resource, the depth of integration into a given program or course, and who will miss it if it is cancelled are generally relegated to the realm of anecdotal evidence. Researchers at Nevada State College have made efforts to remedy this gap in knowledge by analyzing EZProxy logs, which can be set up to capture unique user identifiers at the point of authentication into library electronic resources. When matched with institutional data about those users, this data can answer questions such as cost per user, percent of users within the institution as a whole or a specific department, and frequency and depth of use for the average user. Ultimately, librarians can even discover if library users have higher GPAs or make better grades on research assignments. This paper will provide a summary of the work of the research team at Nevada State College, currently in the midst of their first year of EZProxy log analysis.

About Nevada State College

Nevada State College (NSC) was founded in 2002 and quickly became the fastest growing institution of higher education in the state of Nevada, and one of the fastest growing in the country. The College has a diverse student population of about 3,400 students. Over fifty percent of the students are from culturally diverse backgrounds. More than sixty percent of the students are first-generation college students, and about seventy-five percent are women (Nevada State College 2015).

Serving a student population traditionally underserved by higher education is important to the mission and vision of Nevada State College. The College is constantly searching for solutions to help students succeed in their educational goals. One centerpiece of this effort is the use of data to drive decision-making. The Office of Institutional Research makes use of several online data dashboards and a behind-the-scenes predictive analytics framework to identify indicators that a student may be struggling, and proactively recommend services that positively correlate with student success outcomes such as advising and tutoring.

The Marydean Martin Library at Nevada State College initially became interested in EZProxy log analysis as a way of contributing to this predictive analytics framework by becoming a recommended

service for student success. This paper will focus on the collection development aspects of the data. However, using the data to assess the library's contribution to student success and to identify and address equity gaps in library use are also important goals of the project.

Nevada State College's Marydean Martin Library is the first digital library in the state of Nevada. There are no open book stacks in the library building, and the vast majority of resource use comes from the online collection encompassing more than 1.4 million e-books and roughly 5 million scholarly articles.

How the Data Is Collected

EZProxy is OCLC's authentication system for managing access to online library resources such as e-books and electronic databases. Many academic librarians know it as the off-campus authentication system, but at Nevada State College EZProxy mediates on- and off-campus access. Librarians at Nevada State College made the choice to prompt for authentication on campus in order to capture all e-resource use in the log files. EZProxy logs are an optional feature of EZProxy, which can be setup to track and save user activity (OCLC 2015). Most libraries do choose to collect some form of a log file for troubleshooting or security purposes, but often the data collected is brief and automatically deleted after thirty to sixty days.

At Nevada State College a comprehensive log file is collected, which includes, most notably, the unique user ID of the individual accessing library resources. The systems analyst in NSC's Office of Information Technology Services (ITS) protects the raw data with encryption. On a monthly basis, she uses a PowerShell script that parses the data into a CSV format and sends it to NSC's Office of Institutional Research (IR) using a scheduled job. In the IR office, the data is uploaded into a SQL database. SQL processes written by IR staff anonymize the data and match it with student data as well as usage data from other services on campus such as the writing center, advising, and tutoring. The library works with IR staff to create and maintain SQL update procedures that cleanup the log files, converting complex fields like the request URL into clear database and platform names whenever possible. Finally, the Director of Institutional Research visualizes the data in an online dashboard using Tableau.

With this workflow, all of the matched data is anonymous and raw library use data is encrypted and protected. So for example, the number of library users with a GPA above 3.5 that accessed the Discovery search is evident from the online

dashboard. However, that a named user with a GPA of 2.8 accessed WorldCat Discovery on October 5th is not stored anywhere in the dataset.

For collection development purposes most of what librarians need to know about resource use is evident from the online dashboard. However, there is sometimes a need for deeper analysis such as the correlation between library use and GPA or the statistical significance of equity gaps in the usage of library resources by ethnicity or gender. When those needs arise, the quantitative Analyst from the Office of Institutional Research has been able to provide the library with statistical analyses based on the anonymous matched dataset.

This project involves a significant collaborative effort across the institution. Some librarians outside of Nevada State College have noted that the barriers to such a partnership on their own campuses may be too high to explore EZProxy log analysis. However, there are still valuable insights to glean from proxy log analysis that do not require any matching with student data. To illustrate this, the first portion of this paper will refer to a dashboard created within the library

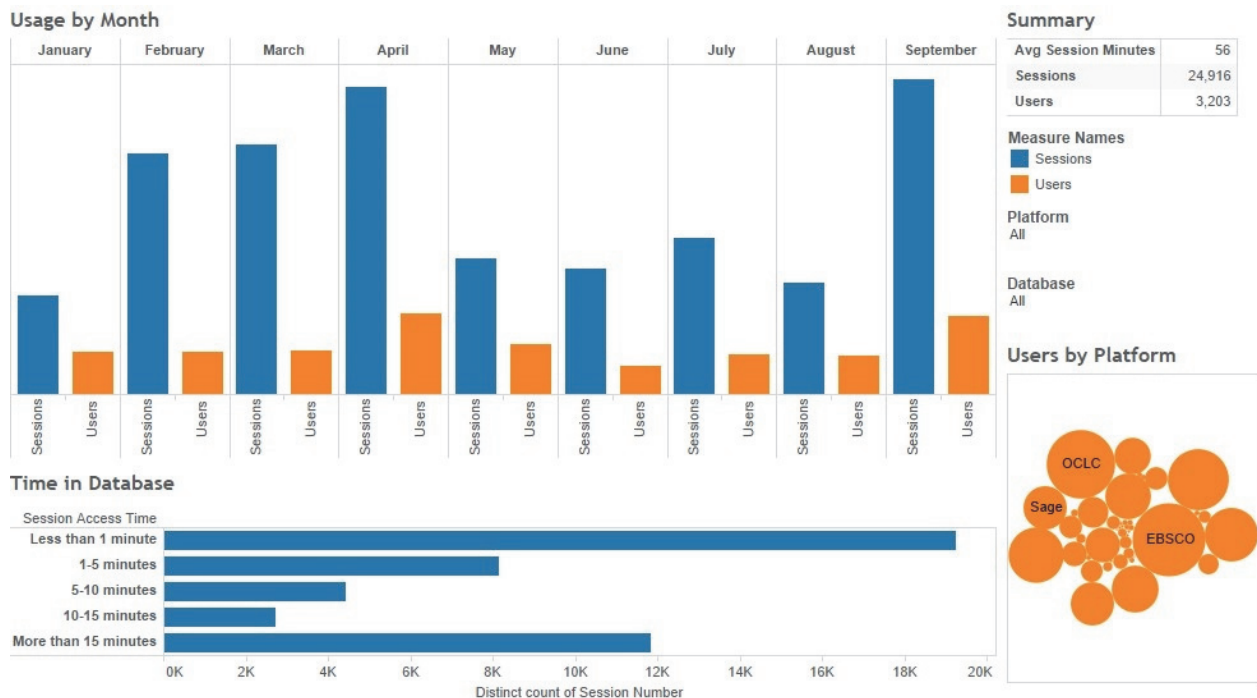


Figure 1. Screenshot from Tableau dashboard of Nevada State College's online resource use from January to September 2015.

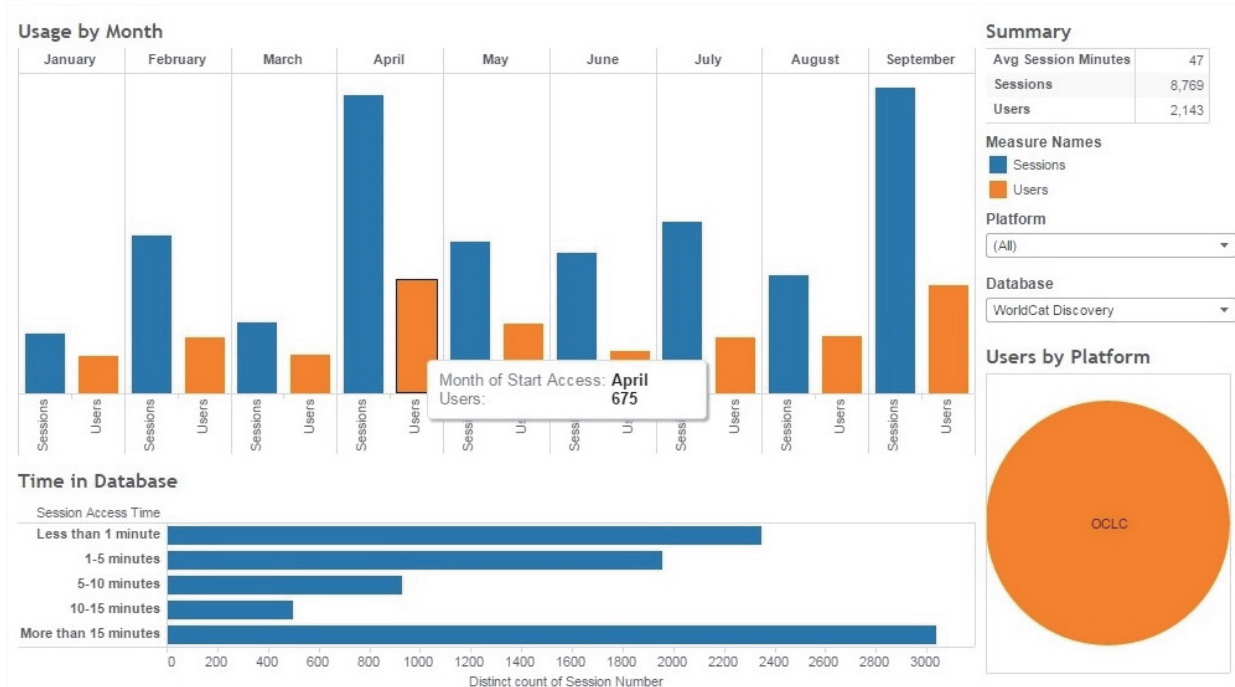


Figure 2. Screenshot of Tableau dashboard limited to WorldCat Discovery use with the cursor over April users bar revealing additional detail.

with anonymized EZProxy log data only. Then, the paper will explore the additional value from matching with student data through the dashboard created from the partnership with NSC’s Office of Institutional Research and Instructional Technology Services.

Early Benefits of the Data

Without matching EZProxy log data with student data, a library can still use the proxy logs to determine the number of unique sessions and users for electronic resources. This can be useful in that it provides a true comparison across all electronic resources, offers an opportunity to calculate return on investment, and gives insight into frequency and patterns of use over time. Figure 1 shows an unfiltered view of an online dashboard created from anonymized EZProxy log data in Tableau. It shows the number of users and sessions by month and overall. A bubble chart displays the number of users by platform indicating high use of EBSCO, OCLC, and Sage.

All of the bars and bubbles in the online dashboard are interactive. Users can roll over them for more detail, or select them to filter the

data presented. For example, Figure 2 shows a view of the dashboard limited to WorldCat Discovery use only using the dropdown menu under database. The cursor is over the users bar for April, showing the number of unique user IDs from the dataset for that month and resource.

Session length and average time are also included, but there are problems with these metrics. Since there is no logout feature of EZProxy, the difference between the start time of the first and last line in the log file for any given session or resource determines access time. The usefulness of this metric varies by platform as some databases create more lines in the log file than others based on similar user actions. User and session counts are the more reliable metrics available from proxy log analysis.

A True Comparison

The most consistent comparison of electronic resource use in the field today is COUNTER compliant usage statistics (COUNTER—counting online usage), but there are two problems with COUNTER. First, not all vendors are COUNTER compliant. Some of the major databases subscribed to by the Nevada State College Library

still do not provide statistics in COUNTER format and a few do not provide usage statistics at all, making consistent comparisons with other resources impossible. Second, there are nuances within the COUNTER standard itself that can make comparing one resource to another challenging. For example, some of NSC's e-book providers only report COUNTER BR1 for title requests. Others only report COUNTER BR2 for section requests. It is clear that counting a use for every e-book title accessed versus counting a use for each section or chapter within a book that is accessed is a different measure, but there is no way of knowing exactly how much section requests are inflated compared to title requests unless the vendor produces both reports. Not all e-book vendors do. User and session counts from EZProxy logs allow librarians to use one standard metric to compare all resources regardless of platform or resource type.

Return on Investment

Another benefit of the EZProxy log data alone is that even without knowing anything about who the users are, having unique user identifiers tied to use can allow for a calculation of number of users and cost per *user*. This metric offers a different perspective than cost per use on whether the library is achieving economies of scale. It also allows for a calculation of return on investment for subscription resources that may be more palatable to administrators than efficiency measures like cost per use. Comparing cost per use to ILL fees or article purchase costs is an excellent measure of efficiency within the library, but it means little to outsiders. By contrast, cost per *user* is comparable to individual subscription fees to calculate how much the library is saving faculty and students by providing access to certain resources. It puts more emphasis on the value of the subscriptions provided by the library, and less on efficient spending of resources allocated to the library.

Usage Patterns

Without matching EZProxy log data with student data, librarians can also get a sense of usage patterns over time. They can calculate new versus repeat users of a resource and begin to

understand how frequently users access certain resources. Currently, NSC's work with analyzing usage patterns from the proxy logs in the early stages. NSC researchers look at the number of users and number of sessions to get a sense of new versus repeat users. For example, for the Discovery search there were 2,143 total users and 8,769 sessions between January and September 2015. Since the Discovery tool was implemented in January, this means that at least 24% (2,143) of the sessions in WorldCat Discovery were from new users and 76% (the remaining 6,626 of the sessions) were from users who had accessed the service before.

Theoretically, a much more advanced analysis should be possible. For example, librarians should be able to determine the number of users who accessed a resource once, 2–5 times, 5–10, and 11 or more. Alternatively, they could analyze how many users accessed a resource once per semester, monthly, weekly, or multiple times per week. If it was of interest, a researcher could even determine what times of day or week experienced the highest traffic in certain resources.

Adding Student Data to the Analysis

The benefits of EZProxy log data alone are interesting, but the true power of EZProxy log analysis is evident from matching this data with data about users. This requires collaboration with campus stakeholders who have access to the data, as well as careful attention to issues such as anonymity and data security. As a result, the Library at Nevada State College chose to collaborate with experts on our campus in the Office of Institutional Research and Information Technology Services to achieve an analysis that goes beyond collection development and begins to address the library's role in student success.

Figure 3 shows a Tableau dashboard created by the NSC's director of institutional research. In this dashboard, the potential of EZProxy log analysis for understanding users is evident. The number of unique users by date from January through May 2015 appears at the top. In the online dashboard, users can adjust the slider below that chart to filter the display of all other dashboard elements to the date range desired. The percent of library

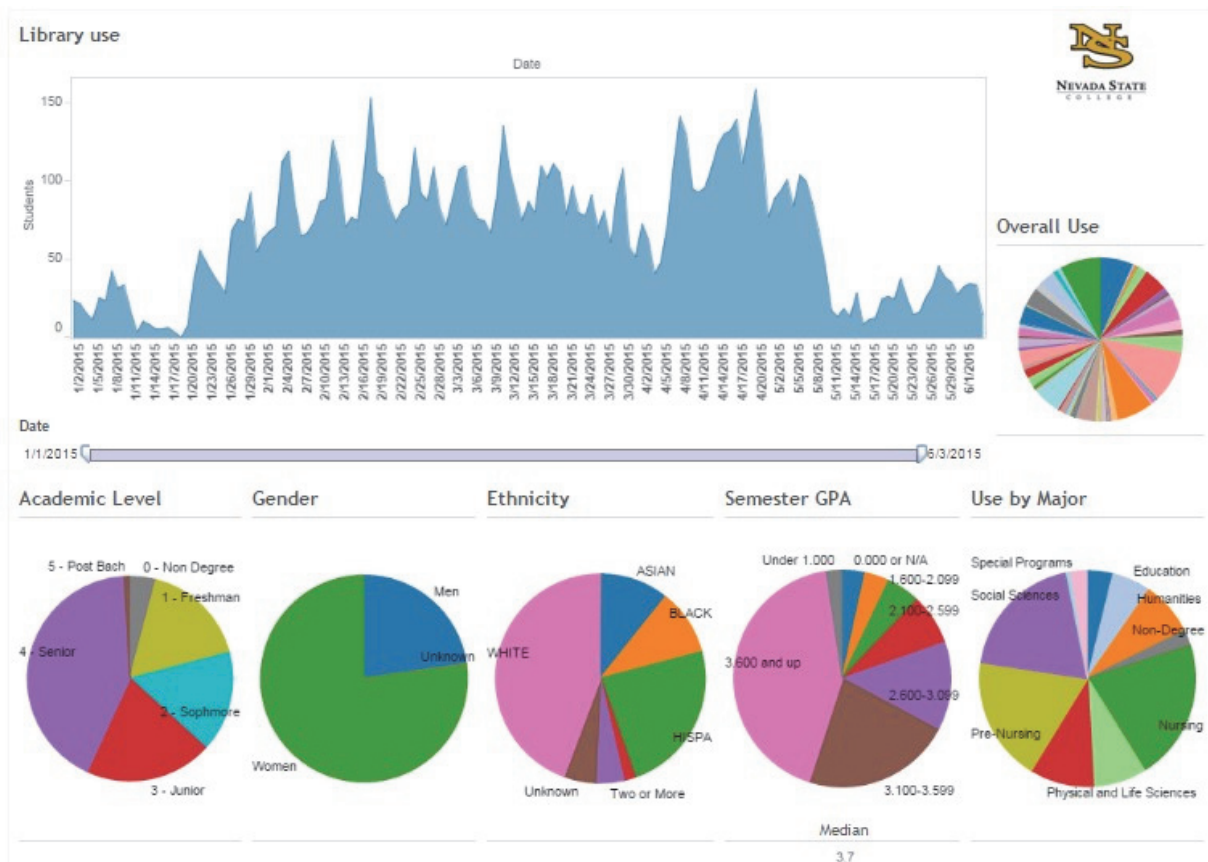


Figure 3. Screenshot from Tableau dashboard of Nevada State College Library’s users.

users at each academic level and in broad disciplinary categories and GPA ranges appears at the bottom. The ethnicity and gender of library users is also included to determine if library users are a reflection of NSC’s diverse student population.

On the right-hand side, there is a pie chart of Overall Use. In the online dashboard, rolling over the slices shows specific library resources such as JSTOR or Academic Search Complete. Clicking a slice will filter the dashboard to usage for that resource only. For example, Figure 4 shows the data limited to use of the WorldCat Discovery search tool.

This dashboard is useful for determining how representative library users are of the Nevada State College students as a whole. For example, underrepresentation from education majors is something librarians at Nevada State College have

already begun to address with outreach to faculty in that program. Filtering to particular resources can offer further insights into the user group for that particular resource. For example, filtering to “historical New York Times” showed librarians at NSC that usage of the resource is broadly interdisciplinary and when considering a switch to the NYTimes.com platform there were several campus stakeholders that needed consultation.

From the dataset behind this dashboard, the quantitative analyst from NSC’s Office of Institutional Research has already begun to communicate to the library about characteristics of library users. For example, her analysis of library users with spring semester data uncovered a statistically significant positive correlation between number of library sessions and GPA. When a full year of data is available for analysis in January 2016, the research team will pursue additional analyses.

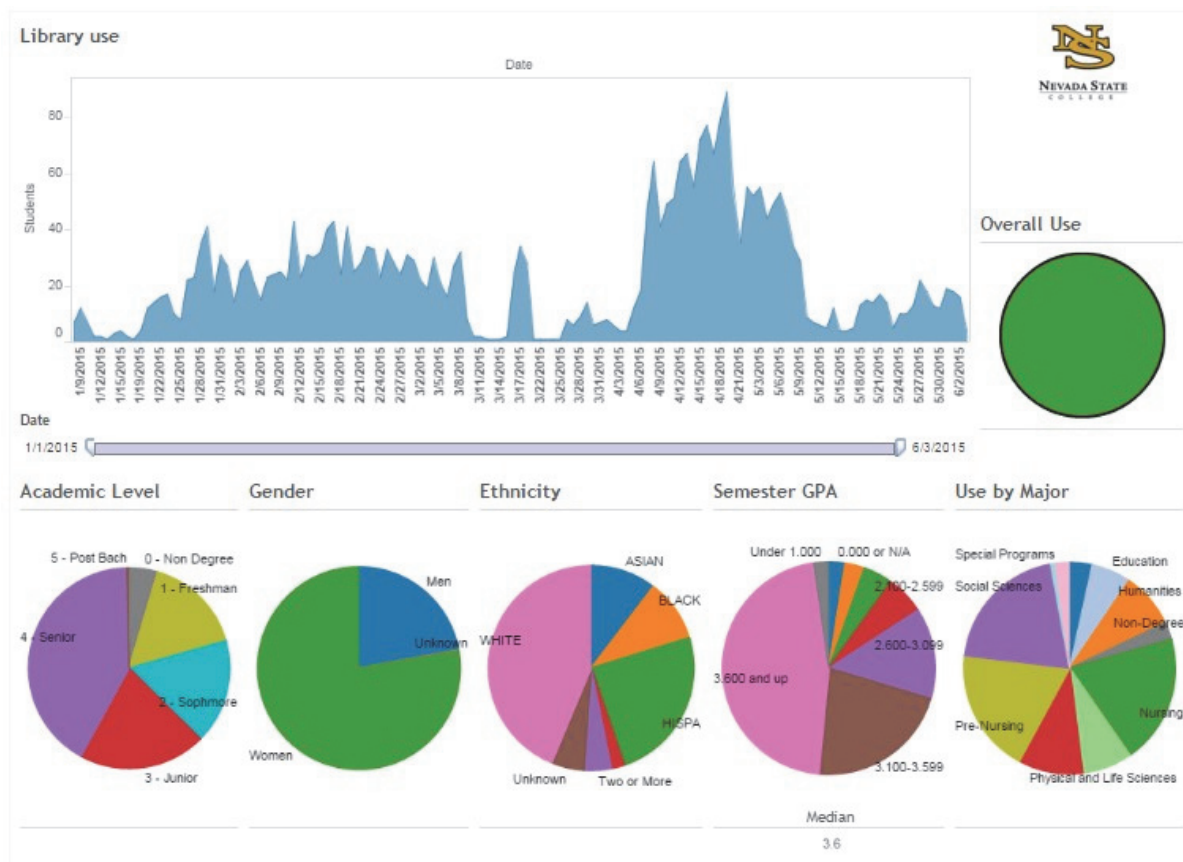


Figure 4. Screenshot of Tableau dashboard of library users limited to WorldCat Discovery use only.

Limitations and Future Directions

There are several limitations of this project, which we are currently attempting to address with further analysis. First, collection use is only one type of library use. Truly understanding the library’s impact on student success requires other data points that correspond to services and spaces. Second, the current statistical analysis is in the exploratory stages. Researchers will need to conduct additional analyses before conclusions that are more meaningful are evident from the dataset. Finally, as with all correlational studies, we cannot infer causation even if we are able to demonstrate statistically significant results.

Incomplete Dataset

As a digital library with on- and off-campus authentication through the EZProxy server, NSC’s EZProxy logs provide a relatively complete picture of collection use. However, there are still pieces missing. There is a small amount of print book circulation from interlibrary loan and other

request services that it is our goal to include in future analyses.

Collection use is also only one small part of what the library has to offer. Librarians at NSC suspect that library services and spaces may be just as or potentially more important to student success than the resources we offer. Several other institutions who have pioneered work in matching student identifiers to student success metrics have combined other types of library use beyond the collection (Cox & Jantti, 2012; Evans, Bowe, Golian-Lui, & Luther 2015; Library Analytics and Metrics Project 2015; Regents of the University of Minnesota). Including additional data points like study room checkouts, library instruction attendance, and research consultations is one of the long-term goals of this project.

Statistical Analysis

So far, researchers at NSC have only run preliminary statistical analyses on the dataset that includes library users only and consists of

correlations that do not attempt to control for potential confounding variables like high school GPA. At the end of one year of data collection, the research team is planning a more advanced analysis that includes success measures beyond GPA like retention and graduation rates, compares users to non-users of library resources, and controls for other variables that may influence college GPA like high school GPA.

Exploring Causation

As with all correlational studies, even if NSC researchers are able to demonstrate a statistically significant positive relationship between library use and student success outcomes, they cannot conclude that library use is the cause of student success. As a result, librarians at Nevada State College are in the early stages of exploring possible experimental designs that may be more suited to inferring a causal relationship. In one pilot project, librarians have worked with an English 101 instructor to add a research assignment to a course that did not have one previously and look for any relationships between library use and grades on that assignment in the fall 2015 semester. The study is a pilot because the sample size is small and there is no control group, but it has allowed researchers at NSC to explore possible methods for implementing an

experiment that might encourage library use through changing course requirements and allow the researchers to look for any significant changes in student learning outcomes as a result.

Conclusion

Despite the limitations, this project has brought the library one step closer to understanding the impact of library resources. Instead of evaluating e-resources on retrievals or search counts, the library can understand the users impacted by those resources. Researchers at Nevada State College are now able to use the dataset to determine if library users are representative of the campus population. If discipline-specific or demographic groups are underrepresented the library can begin to investigate whether services and collections are actually meeting the needs of all students, and if not why not. Finally, the library can demonstrate and work to improve the connection between library use and student success outcomes like GPA. All of these are important steps in shifting assessment efforts from library outputs to library impact.

Acknowledgments

The author would like to thank Tina Buck and Nathaniel King for their comments and feedback.