From Analyzing Abundant Data to Identifying Actionable Steps: A Closer Look at Library Student Data

Mariya Gyendina, Jan Fransen, Carissa Tomlinson, Kate Peterson, and Shane Nackerud
University of Minnesota, USA

Introduction

Student identities impact college experience. A white middle class student speaking North American English likely will have a different experience in their program than a first generation student of color, who grew up in a multilingual household. As we explore the connections between students’ use of libraries and their success in college, it is important to look at the experience of diverse groups of students. In our previous work, we focused our efforts at sharing the results with high level administrators and staff in various departments, including the university libraries. This often took the form of presentation with key findings, for example sharing statistics on the connections between the students’ use of the libraries and GPA or graduation rates. These efforts bolstered our visibility on campus and led us to have a bigger role in large campus efforts such as orientation and first year experience courses. However, the impact of these programs did not reach all students in an equitable way. Our hope is that this small scale analysis will support data-driven targeted efforts. We want to use the data to prioritize targeted outreach interventions. By emphasizing more specific analyses we are able to be more strategic and focused with our outreach. We can take what we know about student experience, department staff and culture, the existing programs, and the staff and create small scale custom interventions such as instruction or outreach efforts. Instead of starting with opportunities as they appear, this method begins with the past data and uses that to proactively seek out opportunities for conversations and learning more about the needs of students in small cohorts.

In addition to continuing our campus-specific efforts, this article builds on the previous work in the field and takes it further by combining statistical analysis and using effect size measures to identify groups for further outreach and new programming. While we focused on models with medium to high effect size, it is important to remember that in education small effect size can still mean large differences. A tiny change in GPA might mean the difference between graduating or having to spend thousands of dollars on another semester of school.

Using large-scale statistical analysis is not a comprehensive research tool for addressing this research question, but building smaller datasets provides a more nuanced picture and can lead to developing more specific, actionable outreach and programmatic efforts. This article describes our approach and lays out a pathway for continuous analytical and programmatic work. While we hoped to pilot a few programs in the fall of 2020, those plans were disrupted by COVID-19 related changes, and we had to put some efforts on hold.

Literature review

After the publication of the Association of College and Research Libraries’ groundbreaking The Value of Academic Libraries: A Comprehensive Research Review and Report in 2010, several academic libraries from around the world have conducted research involving large scale data collection and correlative analysis involving library use and student success measures. Early work in the burgeoning field of library-based learning analytics included pioneering studies from the University of Wollongong in Australia,1 Huddersfield University in the United Kingdom,2 and the University of Minnesota in the United States.3 These early studies were some of the first to collect and analyze large sets of student data combined with
library usage data, focusing high level goals of demonstrating the positive impact library use can have on student success measures such as GPA, retention, and degree completion.

Since 2010 and the publication of these and other early works, the field of library-based learning analytics has grown with other colleges and universities digging deeper and attempting to tie student library usage to more and more specific research questions. In reaction to the University of Minnesota’s 2013 finding that course-integrated library instruction in students’ first year of undergraduate study was negatively correlated with their first-year grade-point average, Gaha, Hinnefeld, and Pellegrino conducted a study that showed an increase in GPA for graduating students who were enrolled in classes that had at least one library instruction session over students enrolled in classes with no library instruction. They suggested that differences in disciplinary grading standards could be the reason for the discrepancy in the two studies, and that grading standards should be taken into account when planning future research in this area. Also in 2018, LeMaistre, Shi, and Thanki published a study that looked specifically at digital resource use at a small teaching college. They found that the correlation between library use and student success is statistically significant with small to medium effect sizes, however they also found that it is not “substantial in magnitude.” This study also paid particularly close attention to effect size in their results, showing that even small effect sizes were valued at their institution. They suggested conducting further research to find greater effect sizes in specific library services.

More recently in 2020, Mayer, Dineen, Rockwell, and Blodgett combined quantitative and qualitative research methodology to answer several research questions to improve the student library experience. Their quantitative research found that library usage in a number of areas is associated with increased persistence and it provided valuable information about students in the aggregate. Their qualitative research created a more holistic view of students and gave the authors a better understanding of how students define success, acquire new knowledge, and find ways to apply their knowledge after graduation. They report that their results “provided us with actionable items to help improve the undergraduate student experience with the library.” Also in 2020, Croxton and Moore shared an important study where the library combined efforts with Academic Affairs, Student Affairs, and other academic and support units to align the gathering and analysis of student engagement data across their units. Together these new partners found “engagement in library, co-curricular, and extracurricular activities has a positive relationship to student success as measured by four-year cumulative GPA and credit-hour completion rates,” while also demonstrating the benefits of creating consistent and transferable data gathering and analysis methods among campus partners.

Of course, this research has not been conducted without controversy or concerns about patron privacy. Maintaining the privacy of library users and the security of library user data has long been of paramount importance in academic libraries, and there has been some criticism of methods in previous studies, both in collection and analysis. In 2018, Jones and Salo argued that there are “moral problems” with learning analytics in that it may involve large data sets of sensitive information essentially created through “student surveillance systems,” which could result in “misleading and harmful paths of action” that may negatively impact students. They further argue that library participation in learning analytics conflicts with at least three of the principles found in the ALA Code of Ethics. In a later study, Jones, et al. further analyzed library participation in learning analytics and resulting publications and found “several gaps in practice” that demonstrate a need for better data management and data security within library projects, as well as better training and collaboration with other disciplines more familiar with this type of research to improve research methods and ethical practices. Finally, Robertshaw and Asher argue that based on their analysis of the results of a number of library-based learning analytics publications, the studies “have not produced findings that justify the loss of privacy and risk borne by students.” Their meta-analysis pays particular attention to the effect sizes reported by the researchers of library learning analytics studies and found that many studies either didn’t report any effect sizes at all, or the effect sizes that were reported were so low (or non-existent) that basing decisions on these findings could be harmful to students.
These studies and the concerns they raised led us to design our research project to specifically focus on exploring the connections between library use and student GPA for smaller groups of students to build a more nuanced understanding of this phenomenon. In this paper we highlight student library usage that has a medium to large effect size and offer possible ways to apply the data to programming and outreach efforts in order to improve the student experience.

**Methods**

This project took an iterative approach to data analysis. In addressing the privacy concerns raised by Jones and Salo,¹ we decided not to collect new data, but instead use a dataset from a previous year (2016–17), since we have no reason to believe that it was significantly different from any other academic year. The dataset combined library data with the data from the Office of Institutional Research and included a number of demographic variables and library use indicators (see Table 1 for descriptions). The students were not identified.

**Table 1. Dataset structure**

<table>
<thead>
<tr>
<th>College</th>
<th>The college in which the student is enrolled. Students who have declared a major also have a department on their record; all students who have not yet declared a major appear as enrolled in the college.</th>
</tr>
</thead>
<tbody>
<tr>
<td>Semester</td>
<td>Fall or spring</td>
</tr>
<tr>
<td>Cumulative GPA</td>
<td>Cumulative GPA including the semester analyzed. For newly enrolled first-year students cumulative GPA is 0.</td>
</tr>
<tr>
<td>Term GPA</td>
<td>Term GPA of a specific semester.</td>
</tr>
<tr>
<td>First generation</td>
<td>Whether the student is first generation. Information is self-reported at the time of application to the university.</td>
</tr>
<tr>
<td>Gender</td>
<td>Students identified as male, female, or unknown.¹</td>
</tr>
<tr>
<td>Ethnicity</td>
<td>Students identified as Asian, Black, Hispanic, Pacific American, White, Unknown.</td>
</tr>
<tr>
<td>International student</td>
<td>Students identified as international, not international (domestic), unknown.</td>
</tr>
</tbody>
</table>

¹ This is not representative of how our group thinks about gender, but we had to use the data provided by the university.
<table>
<thead>
<tr>
<th>Age</th>
<th>Students identified as being in one of the age brackets: 18 and under, 19–20, 21–24, 25–35, 35 and up.</th>
</tr>
</thead>
<tbody>
<tr>
<td>New high school</td>
<td>Identifies students who graduated high school in the semester prior to enrollment.</td>
</tr>
<tr>
<td>Full/part time</td>
<td>Identifies students as part time or full-time based on university definitions: full time students take 12+ credits as undergraduate, 6+ as graduate.</td>
</tr>
<tr>
<td>University status</td>
<td>Undergraduate, graduate, professional, fellow.</td>
</tr>
<tr>
<td>Used instruction</td>
<td>Student was enrolled in at least one course that had an instructional session with a Libraries staff.</td>
</tr>
<tr>
<td>Used reference</td>
<td>Student had at least one interaction with reference staff.</td>
</tr>
<tr>
<td>Used intervention</td>
<td>Student either used reference or used instruction</td>
</tr>
<tr>
<td>Used reserves</td>
<td>Student used course reserves.</td>
</tr>
<tr>
<td>Used digital</td>
<td>Student used a digital resource.</td>
</tr>
<tr>
<td>Used circulation</td>
<td>Student checked out a book/other physical material.</td>
</tr>
<tr>
<td>Used computer</td>
<td>Student logged into a public computer located in a library.</td>
</tr>
<tr>
<td>Used resource</td>
<td>Student used one or more of reserves, digital, circulation, or computer.</td>
</tr>
<tr>
<td>Used libraries</td>
<td>Student used any captured aspect of the library (combines all library use indicators).</td>
</tr>
</tbody>
</table>

For this round of analysis, we limited the analysis to the five colleges with the highest undergraduate enrollment: College of Liberal Arts, College of Science and Engineering, College of Biological Sciences, College of Education and Human Development, and College of Food, Agricultural and Natural Resource Sciences. Initially, we developed models with a minimum of 200 hundred students (100 students who had a particular type of library interaction, and 100 students who did not have one). However, we quickly found out that in this analysis we were missing many groups, especially minority ones, due to being a predominantly white institution. Therefore, we lowered the number of students to 50 (25 yes, 25 no).
With so many potential combinations of variables to take into account, we developed a VBA script to generate both source files for each college and the R code to run the statistical analysis. This method also supports reproducibility, as we can run the R source code later to confirm results or using source data from a different academic year.

The VBA script iterated through the demographic variables and generated the code necessary to create a subset of records for students within each combination of variables and perform the necessary statistical calculations. For example, a subset might be engineering students identifying as female who are not first generation college students and are classified as international. For each subset and use indicator, the R code:

- Checks to see that there are enough people who did and did not participate in the library use indicator to generate a sample (there must be at least 25 yes and 25 no)
- Generates a sample with 25 students on each side (did and did not participate in the use indicator)
- Uses both Term GPA and Cumulative GPA to perform a Levene Test and calculate p-value, Cohen’s f, and eta-squared
- Writes the results to a spreadsheet

Once the R code ran, we used the spreadsheet to evaluate the results.

If the models satisfied the statistical requirements for validity (the Levene Test), we explored further, looking at the p-value, Cohen’s f, and eta-squared. The p-value shows the likelihood of the model describing a random correlation and is set at 0.05. Cohen’s f is a measure of effect size, with the following ranges: 0.1–0.25 small effect size, 0.25–0.4 medium, and >0.4 large. Eta-squared is another measure of effect size, looking specifically at the proportion of variance in the dependent variable explained by the independent variables.

**Data**

After the final round of coding, we spent time as a group looking at the available models and discussing different ways to approach them. After considering several options, such as focusing on models with the highest effect size measures or the ones with the largest difference between the number of student who interacted with the library and who did not, we decided to focus on two directions: looking at the models focusing on interactions that are somewhat within our control, such as instruction, and looking at the models describing the experience of students from minority and underrepresented groups, which reflects the broader organizational focus on equity and inclusion.

Below we present a brief discussion of several models representing different colleges and student groups, combining statistical analysis with potential actionable steps. The models focus on instruction as library interaction type, because instruction-focused models showed moderate to high effect sizes, had an almost even split between students who had and did not have the interaction, and leaves an opportunity for the libraries’ staff to explore opportunities for increasing the outreach to the instructors and college administrators.

**Results**

The analysis returned 407 models showing the connections between library instruction and positive difference in students GPA with moderate to high effect sizes. In this paper we are highlighting seven cases that we identified as having the highest potential for high impact interventions. After reviewing the statistical models we identified the following cases as the ones where targeted interventions would have the most impact. We tried to look across colleges and demographic groups, while keeping in mind
capacity and logistical constraints. Table 2 presents the summary of cases followed by a more detailed description.

The list of possible programming and outreach opportunities represents our brainstorming and is not meant as a comprehensive list. We tried to include various stakeholders in the university community: administrators, advisors, instructors, and students. However, instruction might not be fully within the students’ control, so we are mindful of the need to be careful in approaching the student groups with this topic.

Table 2. Summary of cases.

<table>
<thead>
<tr>
<th>College</th>
<th>Group</th>
<th>Total number</th>
<th>Had instr.</th>
<th>Did not have instr.</th>
<th>P value</th>
<th>Cohen’s f</th>
<th>Eta sq.</th>
</tr>
</thead>
<tbody>
<tr>
<td>CLA</td>
<td>First generation</td>
<td>3562</td>
<td>562</td>
<td>3000</td>
<td>0.44</td>
<td>0.298*</td>
<td>0.08</td>
</tr>
<tr>
<td>All</td>
<td>Black male</td>
<td>551</td>
<td>56</td>
<td>495</td>
<td>0.14</td>
<td>0.37*</td>
<td>0.12</td>
</tr>
<tr>
<td>CEHD</td>
<td>Female first generation</td>
<td>550</td>
<td>102</td>
<td>448</td>
<td>0.18</td>
<td>0.354*</td>
<td>0.11</td>
</tr>
<tr>
<td>CFANS</td>
<td>Full-time</td>
<td>1632</td>
<td>283</td>
<td>1349</td>
<td>&lt;0.01</td>
<td>0.549**</td>
<td>0.232</td>
</tr>
<tr>
<td>CLA</td>
<td>Female Native American</td>
<td>122</td>
<td>29</td>
<td>93</td>
<td>0.006</td>
<td>0.419**</td>
<td>0.15</td>
</tr>
<tr>
<td>CSE</td>
<td>International first generation</td>
<td>156</td>
<td>35</td>
<td>121</td>
<td>0.005</td>
<td>0.424**</td>
<td>0.152</td>
</tr>
<tr>
<td>CBS</td>
<td>Asian first generation</td>
<td>94</td>
<td>33</td>
<td>61</td>
<td>0.003</td>
<td>0.448**</td>
<td>0.167</td>
</tr>
</tbody>
</table>

* medium effect size
** high effect size

Case 1: First generation in College of Liberal Arts

The College of Liberal Arts is by far the biggest undergraduate college on campus with over 13,000 students. By using the methods above, we found that for the 3,562 first generation students in CLA in spring semester of 2017, 562 of them received library instruction and 3000 did not. The Cohen’s f is 0.298, indicating a medium effect size. Eta-squared is .08 which indicates that having library instruction explains 8% of the difference in GPA between students who had library instruction and students who did not.

As we continue working with this population, we can then take a look at strategic ways to increase the proportion of students who experience library instruction by identifying the courses students are likely to take and working towards curriculum integration. We also plan to meet with the first generation student
group and President’s Emergent Scholars to learn more about students, their experience and ideas. Finally, we plan to reach out to the academic advisors in CLA to learn more about students’ experience.

**Case 2: Black, male students**

In the spring of 2017, there were a total of 551 black, male, students enrolled across the five major colleges. Of this group 56 (about 10%) received library instruction and 495 did not. The Cohen’s f for this analysis 0.37, which shows medium effect size, while eta-squared is 0.12, indicating that 12% of the difference in GPA between students who had and did not have library instruction can be explained by this variable.

As we consider outreach to this group of students we plan to start by gathering more data through talking with students in the black student union and the coordinator of African American recruitment. Additionally, we hope to reach out to the orientation staff, so we can include targeted materials in the orientation programming.

**Case 3: Female first generation students in CEHD**

The College of Education and Human Development enrolls approximately 2,450 students, and in the fall of 2016 there were 550 female first generation students; 102 of them received library instruction and 448 did not. The Cohen’s f for this model is 0.354, which indicates medium effect size and the eta-squared is 0.11, showing that 11% of the GPA gap between the students who received and did not receive library instruction can be explained by this factor.

The priority outreach efforts in this case would target better integration with the first year inquiry course that all incoming students are required to take, as well as reaching out to the TRIO center. We also plan to apply for a microgrant from the First Generation Institute to develop a program specifically addressing the needs of this group of students.

**Case 4: Full-time students in CFANS**

The College of Food, Agriculture and Natural Resource Sciences (CFANS) enrolls 2,000 undergraduate students, with 1,632 full-time students. In the fall of 2016, 283 students received library instruction, and 1,349 did not. The effect size of this model is 0.549, which is high, and the eta-squared is 0.232, indicating that 23.2% of the difference in GPA between the students who received and did not receive library instruction could be explained by this variable.

Since this model presents college-level data, we plan to focus on two larger scale initiatives by collaborating with the Writing Enriched Curriculum program and presenting these findings to program directors to encourage further curriculum integration.

**Case 5: Female Native American students in CLA**

As mentioned above, CLA enrolls more than 13,000 students, including 122 Native American female students, 29 of whom received library instruction (93 did not). This model has Cohen’s f of 0.419, which represents high effect size, and the eta-squared is 0.15, indicating that 15% of the difference in GPA between the students who received and did not receive library instruction can be explained by this variable.

In identifying the outreach opportunities for working with this group we plan to start with bringing these data to the CLA academic advisers and analyzing which courses these students are more likely to take, so we can target instructional opportunities.
Case 6: International first generation students in CSE

The College of Science and Engineering enrolls approximately 5,400 students, and in the fall of 2016 156 international students were first generation. Within this group, 35 students had library instruction and 121 did not. The Cohen’s $f$ for this model is 0.424 (high effect size), and the eta-squared is 0.152, indicating that 15% of the difference in GPA can be explained by the library use variable.

This is a relatively small group of students that can be overlooked by programs looking to support international students and first generation students, thus, we plan to collaborate with both units to develop awareness and provide targeted opportunities.

Case 7: Asian first generation students in CBS

The College of Biological Sciences enrolled 2,212 students in the fall of 2016. There were 94 Asian first generation students in that group, 33 of whom had library instruction as part of their courses (61 did not). The Cohen’s $f$ for this model is 0.448, showing high effect size, and eta-squared is 0.167 demonstrating that 16.7% of the difference in student GPA can be explained by this factor.

In this case we plan to focus on curriculum integration through partnering with the Writing Enriched Curriculum program and reaching out to the Asian Pacific American Resource center to identify collaboration opportunities.

Actionable steps

The reporting of library impact on students is important, but it is really just a starting point. In order to have real impact, we must rely on that data to enact change. The data has shown us that there are areas where the libraries have significant impact on specific demographics and in some cases, those demographic groups have minimal interaction with the library. With this data in hand, the group is working to target various levels of the organization to influence change in outreach and instruction priorities.

Because our organization is large and multilayered, our approach has been to bring this information to both the administrators and the staff in hopes to change libraries-wide priorities as well as individual priorities. At the time of writing, we have met with the Research and Learning directors and we are beginning to meet with staff at department meetings in the Research and Learning division. Our long term goals for these meetings are to encourage individuals and departments to prioritize work that has the greatest impact and consider new ways to support this work such as by developing new teams or groups to focus on this work. For individuals, especially liaisons, we are hoping to spur targeted instruction or outreach. Our shorter term goals are to increase awareness of the project and improve data collection.

For the department meetings, our strategy is to bring subject specific data as well as college specific strategies. For example, in meeting with the libraries departments that serve the College of Liberal Arts (CLA), we were able to show that only 23 of the 122 Native American women in the College of Liberal Arts had received libraries instruction. At the same time, we are able to show that library instruction is connected with student success. In this example, CLA liaisons may want to consider outreach to Native American women by targeting student organizations or further analyzing data to see which courses Native American women are more likely to take and then target those classes for library instruction. Another example is that the data shows us that international, first generation students in the College of Science and Engineering have a low likelihood of receiving library instruction (35 of the total 156 students). Liaisons to departments in this college may want to hold targeted events or workshops in collaboration with the International Student and Scholar Services department.
For meetings with administrators, the hope is to spark larger prioritization of this work in our institution. This is particularly important this year as our organization develops a new strategic plan. Buy-in from administrators, particularly department directors, is also important when asking individual liaisons to help with this work. In an organization as large and complex as ours, much of our actionable work will be focused on spreading the word and creating buy-in throughout the organization. This work will take persistence and patience, especially at a time when many people have reduced capacity due to the pandemic.

**Conclusion**

The goal of this project of course is not simply to increase numbers. Like so many staff and faculty on campus, we seek to get a sense of the individual students behind these numbers and how we can support and teach them when they need inquiry and research skills. We want to make sure we are not missing chances for students to learn to use the libraries and make sure there are fewer times when students are less successful because they do not have the skills needed to effectively use the scholarly resources from the libraries.

We designed and implemented this project guided by the understanding that diverse groups of students experience college in different ways and potentially benefit from a variety of support models. This article presents our efforts to build a richer picture and identify strategic, thoughtful and long term ways to support student learning and success, using measures from statistical analyses to facilitate selecting priorities. We believe that all of these ideas have the potential to lead to positive outcomes; however, given the current reduced capacity we hope that this data will support our efforts to make informed choices about where to focus our efforts.

This project, as all large-scale quantitative projects, provides a broad picture of the phenomenon and misses the rich details and understanding of the context. It is by no means comprehensive in its analysis. We would like to see it continued with multiple small-scale qualitative studies investigating the details of connections between libraries instruction and student outcomes as well as the impact on students from different groups.

—Copyright 2021 Mariya Gyendina, Jan Fransen, Carissa Tomlinson, Kate Peterson, and Shane Nackerud

**Author Biographies**

**Mariya Gyendina**

Mariya Gyendina is a Learning and Inclusion Strategist for the UMN Libraries. She is also working on her dissertation about the experiences of students in online writing centers. Mariya's research interests focus on language acquisition, assessment, and best practices in instruction.

**Jan Fransen**

Jan Fransen is the Service Lead for Research Information Management Systems at University of Minnesota Libraries. In that role, she works across divisions and with campus partners to provide library systems that save researchers', students', and administrators' time and highlight the societal and technological impacts of the University's research.

**Carissa Tomlinson**

Carissa Tomlinson is the Physical Sciences & Engineering Director for Research and Learning at the University of Minnesota Libraries. She earned her MLIS from Dominican University in 2007 and started
her career at a residency program at Vanderbilt’s biomedical library. She then spent 11 years at Towson University in a variety of roles related to information literacy, health sciences, student engagement, and management.

**Kate Peterson**

Kate Peterson is the Undergraduate Services librarian at the University of Minnesota-Twin Cities. She is the librarian for Writing Studies, First Year Writing, the University Honors Program and supports Orientation and Undergraduate services.

**Shane Nackerud**

Shane Nackerud has worked at the University of Minnesota - Twin Cities for 22 years, currently as the Co-Interim Director of Content Services and Technology Lead of Library Initiatives. Shane's research interests include library use assessment, libraries and e-learning, affordable content integration, and academic publishing. Shane received a B.A in history from Augustana College in 1994 and his MLS from Indiana University in 1995.

**Endnotes**

References


http://dx.doi.org/10.4135/9781412961288.
