

# Correlation Between Library Instruction and Student Retention: Methods and Implications

Mary K. O’Kelly, Jon Jeffryes, Maya Hobscheid, and Rachael Passarelli

Eight years of data from Grand Valley State University Libraries show a positive correlation between in-class library instruction and student reenrollment the following fall semester. Using consistent statistical methods over time, controlling for confounding factors, and using a large population (N>16,000 annually) and strict protocols that restrict library employees from having access to individual student data, researchers have developed a replicable, secure process for exploring the relationship between library engagement and student success that prioritizes data security and subject anonymity alongside scalability. Results have informed modifications to the Libraries’ instruction program and have raised interesting additional questions about the impact of having classroom faculty who actively engage their students with the academic library.

## Introduction

Universities request their libraries to demonstrate the impact of their work with concrete data drawn from a variety of assessment methods in order to make decisions about resource allocation and academic support programming. In response to this kind of evidence-based decision making, libraries have begun looking at the impact of a variety of services and resources on student success metrics, such as grade point average (GPA) and student retention. This paper describes a reproducible methodology and findings that other institutions may adopt to demonstrate the impact of their course-integrated library instruction and to inform further research into the relationship between library engagement and academic success.

Grand Valley State University (GVSU) Libraries found a statistically significant positive correlation between course-integrated, in-person, librarian-led instruction and student retention every year between 2012 and 2020.<sup>1</sup> These findings have been methodologically stable and replicated for eight academic years. This paper shares the unique methods—including analysis scripts—and the results of the analysis for validation and reuse throughout the library analytics community. The paper also provides an overview of the findings and how the results have

---

\* Mary K. O’Kelly is Associate Dean for Education and User Services at Western Michigan University, email: [mary.okelly@wmich.edu](mailto:mary.okelly@wmich.edu); Jon Jeffryes is Associate Dean of Curriculum, Research, and User Services at Grand Valley State University, email: [jeffryjo@gvsu.edu](mailto:jeffryjo@gvsu.edu); Maya Hobscheid is Instructional Design Librarian at Grand Valley State University, email: [hobschem@gvsu.edu](mailto:hobschem@gvsu.edu); Rachael Passarelli is Research Analyst at Grand Valley State University, email: [passarer@gvsu.edu](mailto:passarer@gvsu.edu). ©2023 Mary K. O’Kelly, Jon Jeffryes, Maya Hobscheid, and Rachael Passarelli, Attribution-NonCommercial (<https://creativecommons.org/licenses/by-nc/4.0/>) CC BY-NC.

been used to inform and iterate the Libraries' instruction program in order to give a proof of concept on the instructional improvements such studies can facilitate.

The methodology used to determine this positive relationship between library instruction and student retention is made even more significant by the strong adherence to student data security and subject anonymity followed throughout the study. The datasets were pulled from well-controlled course registration data collected centrally by the university and then combined with individual library instruction session data collected by instruction librarians. University Libraries and the university's Institutional Analysis (IA) department worked together to connect library instruction with fall enrollment data and then analyzed that data annually. This methodology prioritized security and anonymity by retaining all private institutional data within the secure environment of IA; the Libraries never held any personally identifying student data.

### *Review of the Literature*

This study is grounded in a body of literature that explores the specific connections between different aspects of library interaction and student retention. These studies followed earlier, broader research on the role of libraries and student engagement, such as that of Kuh and Gonyea, which posited that "Another research question is whether student use of the library and interactions with librarians are associated with persistence and graduation, net of other factors."<sup>2</sup>

A few years later, Oakleaf's *Value of Academic Libraries: A Comprehensive Research Review and Report*<sup>3</sup> helped spark a variety of new research approaches for measuring and communicating the impact of academic libraries. That report focused on "library value *within the context of overarching institutions*."<sup>4</sup> One recommended approach to ensuring institutional connections was through collaboration with institutional data partners so as to make it possible to connect library data with university data: "Integrating library data with institutional data is critical; without joined data, joint analysis is difficult."<sup>5</sup> This recommendation became a foundational approach for this study.

Other researchers have adopted this approach of comparing institutional-level existing data with descriptive data of library activities, and a burgeoning genre of impact research has analyzed student academic data in connection with library services, tools, and expertise. Those studies range from investigating libraries' participation in recruitment to identifying more specific markers of libraries' impact on retention.<sup>6</sup> Such retention-focused papers have described relationships between student grade point averages and retention with the use of library collections;<sup>7</sup> student retention with general library use;<sup>8</sup> library, instruction, and academic support expenses per full-time equivalent (FTE) population with retention;<sup>9</sup> staffing per FTE with retention;<sup>10</sup> library expenditures with retention;<sup>11</sup> student use of library resources with self-reported development on learning outcomes beyond information literacy such as critical thinking, writing skills, and reading comprehension;<sup>12</sup> and book borrowing and e-resource use with the level of degree results.<sup>13</sup> More comprehensive collections of this type of study can be found in Oliveira's<sup>14</sup> review article on retention studies and the meta-analysis conducted by Robertshaw and Asher<sup>15</sup> on library learning analytic studies.

### **Background**

For eight consecutive years, GVSU, a masters-large comprehensive university with approximately 24,000 students, has consistently identified a statistically significant difference

in year-to-year retention between students who had a librarian visit their class compared to students who did not.

The academic year at GVSU is divided into three sections: fall semester (August to December), winter semester (January – April), and spring/summer session (May – July). This study defines student retention as reenrollment for the subsequent fall semester.

### *Relationship with the University's Institutional Analysis Office*

The first step to implementing a cross-campus collaborative analytics study was to gather administrative support. The Libraries' former head of instructional services initiated the study by first securing the support of the dean, and then, in partnership with an associate dean of the Libraries, met with the director of IA to explicitly ask for one analyst's time to collaborate on the assessment of library influence on student success.

The pitch to IA outlined the benefits to the university: the partnership would protect student privacy by keeping all student-level enrollment data in the IA office, transmitting only aggregate data to the Libraries. The collaboration would also provide IA with direct access to all library instruction data, and the resulting analysis could be used by any campus assessment department or committee to communicate the university's efforts to measure student success.

An agreement was reached that included the following parameters:

1. University Libraries collects its own data about librarian-led instruction sessions.
2. IA analyzes data in large batches, only once or twice a year, following a predictable schedule (e.g., in the two weeks following the end of the academic year) so that they can plan their workload.
3. University Libraries is responsible for the interpretation of the data (although the analyst would be available to answer questions about statistical methods and results).

The IA department at GVSU manages secure student-level data such as GPA, enrollment, and year in school, which was needed for the correlation analysis. It was paramount that this data remained secure, and therefore it was never transferred to the Libraries. Instead, descriptive library instruction data (which contained no student-level data; librarians do not take attendance) was transferred to IA in late August or early September every year for the annual statistical analysis. This close partnership allowed the Libraries to correlate instruction with student enrollment without jeopardizing student privacy and permitted the Libraries to leverage expert analysis skills that otherwise were not available internally. This approach also proactively bypasses the common challenges of campus data silos, which can impede analytics studies.<sup>16</sup>

After the agreement was made and an analyst assigned, the head of instructional services met with the analyst to provide the context of the project, explaining the research questions to be answered and discussing the intent to connect the findings to the institution's strategic plan. They also communicated with the campus institutional review board (IRB) office to confirm that this anonymized analysis of existing educational records and normal library instruction session descriptive data was exempt from IRB review and that no identifying information was to be given to the researchers in alignment with FERPA guidelines. As discussions between the Libraries and IA proceeded, a list of twenty-four initial inquiry questions was developed by carefully exploring what the librarians wanted to know about the scope and reach of their instruction program and, of those initial queries, which ones were realistic for the analyst to pursue (e.g., How many students did the instruction program reach? What percentage of

freshmen participated in a library session? Which academic programs did the librarians not reach?). With iteration and input from the instruction librarians and library administration, that list eventually grew to thirty questions (see Appendix A). Those assessment questions framed and helped focus the entire scope of the study.

Even though only one of those questions emerged as the foundation of the current study's hypothesis, University Libraries sends the entire list of questions to Institutional Analysis annually for three reasons: to ensure consistent longitudinal analysis, to have ready answers about all instruction trends (even those not related to retention), and—in a respectful nod to the pressures on a university analyst's time—to provide IA with a very well-structured, predictable request that can reliably be used to plan departmental workload. After spending time to set up the SAS data analyzation script in SAS Enterprise Guide 7.1 analysis software, the analyst expects to spend no more than four hours per year on calculating the data and creating the final aggregated Excel report that will be sent to the Libraries.

### *A Note about Privacy*

As mentioned above, maintaining the security and confidentiality of university-held enrollment records was intentionally included in the original design of this study. The university collects student-level data such as major, GPA, and enrollment and holds those secure, and IA staff handled all student data for the purposes of this study. They have the expertise, software, and data controls to ensure it remained secure. IA never shares that individual data with the Libraries. Instruction librarians enter in all their own session-specific instruction data, but they do not take attendance or collect student names or identification numbers. The enrollment data is used to evaluate a regular educational program in typical educational settings taught by faculty librarians, and no student is denied any instruction nor experiences any altered instruction in this methodology.

Literature in the field of library and information science reflects concerns about ethics and privacy of student data.<sup>17</sup> As outlined in the methodology, student privacy issues were considered seriously in the planning of this study from the outset. This methodology describes an approach to assessing library instructional activities with significant measures taken to protect student data. Briney<sup>18</sup> cited O'Kelly's particular methodology (as presented previously) as an example of an analytics study that "handled data well" through a "neat avoidance of the burden of security by having the library hold only aggregate data." The Libraries' interest lies only in the broader trends that can be found in aggregate data to make iterative improvements to the instruction program and communicate the impact of library instruction to campus stakeholders.

### **Methods**

The population studied is drawn from the total enrolled student population at GVSU, and students who enrolled in any course section at any level that included library instruction are compared to students who did not during one academic year. The analyzed population is limited to those students who encountered library instruction through a credit-bearing course. Those students had to have at least one course in their schedule that had at least one section experience a library instruction session, even if it wasn't that student's section. In other words, if a student had zero opportunity to encounter a librarian through any course because no sections of any course in their schedule included library instruction, their data was stricken

from the analysis; this explains why the N of the population analyzed differs from total annual university enrollment.

By limiting the analysis to in-class librarian presentations and not including any voluntary workshops or orientations, this study removes the complicating variable of student agency and counters any unintended bias that may occur through over-representation of intrinsically motivated students who sought out library instruction on their own; this is a direct response to other studies that only investigate self-motivated, voluntary student engagement with library services and collections.

University Libraries' instruction librarians recorded all library instruction data into Springshare's LibAnalytics (now LibInsight) datasets. The Institutional Analysis department uses Ellucian's Banner program, an enterprise resource planning and student-information system, to manage student data.

The Libraries' instruction datasets used three fields to collect standardized information about the courses in order to cleanly match these three fields with corresponding Banner fields: course code, course number, and section number. In addition, librarians recorded their own name and the name of the course's lead instructor, to assist with cleaning data; in cases when it was not clear whether the section and course number were accurate, the professor name could be used to cross-check for accuracy. Other fields are date of instruction, location, duration (in minutes), and content of instruction session. Librarians did not take attendance because, according to the IA department's experience with similar enrollment analysis, typical student absences fall within the margin of error. For the purposes of this study, it is assumed that a typical absence pattern occurs in course sections with librarian instruction.

The types of courses taught include a wide and diverse variety of disciplines. As librarians provide support to all First-Year Writing instructors, the majority of courses reached are in the writing department. There is also a high number of courses reached in STEM and health science departments. The instruction program usually reaches a higher percentage of freshmen than of other classes, though the percentage of students reached in each class varies from year to year. While each librarian develops their own instructional content in collaboration with their disciplinary faculty, the Libraries used a set of information literacy core competencies as guiding learning outcomes for the first five years of this study and now use an information literacy assessment rubric.

At the end of each academic year (early August), the librarian responsible for instruction assessment downloads the annual instruction data, aggregated from all the entries of individual instruction librarians, from LibAnalytics/LibInsight into Excel, cleans the data to remove extraneous fields of information irrelevant to the current study, and ensures all entries are consistent. The clean spreadsheet is then forwarded by the Libraries to the analyst in IA. These data are accompanied by the primary research hypothesis and a secondary list of additional assessment questions.

The analyst compares the Libraries' instruction data to Banner enrollment data for the entire population of enrolled students using secure computers in the IA office. The analyst matches the Libraries' instruction datasets to Banner data using the course section and academic term fields. (The analyst determines the term with a simple IF function in Excel, using the date of instruction to logically determine academic term.) With the term and course information matched, IA then creates a subset of student identification numbers for those students enrolled in the courses with library instruction.

To break it down further, the IA analyst follows these four procedures: determine which courses had at least one section with a library instruction session; then determine which students were enrolled in the sections that had a library instruction session and which students were not; further determine which students reenrolled the following fall semester; and finally compare those populations of students. After the correlation analysis, results are returned to the Libraries in aggregate.

Methods were chosen for their ability to answer the central research question (R1) and potentially reject the null hypothesis ( $H_0$ ).

R1 – *Of the students who saw a librarian in class, what percentage of them reenrolled for the following fall semester compared to students who did not see a librarian?*

$H_0$  – *There is no relationship between library instruction and student retention.*

This question emerged as the most critical of the thirty assessment questions GVSU Libraries had selected to measure and evaluate the instruction program, as retention has proven to be a key goal for the institution. The remaining questions ranged from simple queries about how many students were reached to very complex questions about the intensity effect on reenrollment for students who saw a librarian in class multiple times.

### ***Statistical Analysis***

The analyst performed a chi-squared test of independence to determine the relationship between students who experienced an in-person library instruction session and students who reenrolled in the University for the following semester.

A fixed p-value of .05 was used to test significance and eventually tested down to a p-value of .0001. Various confounding factors (ACT score, high school GPA, socioeconomic status, and first-generation status) were controlled using a generalized linear model due to having a mix of categorical and continuous variables.

Odds ratio was used as part of the chi-square test of independence to determine the measure of association. The chi-square test of independence tells whether there is a relationship between the variables studied (students enrolled in classes that include a library instruction session and student retention). The odds ratio indicates the strength of that relationship. The analyst chose odds ratio because of having two binary variables; however, a phi coefficient could also have been used though likely would have returned similar results.

The IA analyst used a script in SAS Enterprise Guide 7.1 to set up the analysis of the data. (See Appendix B for the full script.) The script includes a chi-square test with odds ratio and a generalized linear model.

### **Results**

The following results illustrate the findings at GVSU. The methodology outlined above, employed in collaboration with an Institutional Analysis office, explores the Libraries' instruction program's relationship with student success measures, avoids issues of data siloing, and protects student data security and subject anonymity.

### **Discussion and Implications**

The results of this study suggest that students who participate in courses that had a librarian-led instruction session at some point in the semester have a statistically significant higher chance than their peers of being retained. This study defines student retention as reenrollment for

**TABLE 1**  
**GVSU Correlation between Library Instruction and Student Retention ( $p$ -value = .0001)**

Year	Librarian in Class	No. of Students	Percent Retained	Odds Ratio
2012–2013	No	8,762	70.90	1.14
	Yes	8,763	73.70	
2013–2014	No	9,925	72.10	1.15
	Yes	8,560	74.90	
2014–2015	No	10,040	71.00	1.31
	Yes	8,340	76.10	
2015–2016	No	9,819	71.20	1.23
	Yes	8,799	75.20	
2016–2017	No	10,170	70.49	1.29
	Yes	9,558	75.52	
2017–2018	No	9,347	70.39	1.13
	Yes	8,556	75.35	
2018–2019	No	9,156	70.5	1.2
	Yes	7,979	74.4	
2019–2020	No	9,504	65.7	1.5
	Yes	7,930	74.6	

the subsequent fall semester. These results have been consistent across the length of the study from 2012–2013 to 2019–2020 (Table 1). The significance was tested down to a  $p$ -value of .0001, indicating that the 2.8–8.9 range of percentage differences in retention is statistically significant.

The odds ratio results indicate that students who enrolled in courses that included library instruction sessions do have a higher likelihood of reenrollment than those students enrolled in courses that do not include library instruction sessions, but the magnitude of the difference is small as evidenced by the odds ratio consistently remaining only slightly greater than one. An odds ratio of one indicates there is no association between the two groups. The greater the odds ratio is above one, the greater likelihood of increased occurrence of the event (in this case, library instruction students reenrolling at a higher rate than non-library instruction students). A small difference is an odds ratio less than 1.5 and a large difference is an odds ratio greater than five.

Other studies on student success measures such as retention and GPA are often confounded by student motivation. Motivated students who voluntarily choose to use academic support services, including library services, simply may be more successful students who have higher GPA and retention rates. This study bypasses this limitation by using whole-course-section data. By using the enrollment of an entire course section and focusing the analysis on sections that had a professor engage a librarian in instruction, we have removed intrinsic motivation as a potential confounding factor. It is assumed that when a librarian visits a classroom or a professor brings a class into the library for instruction, those students did not volunteer to seek out that particular form of supplemental academic support.

Using only whole-class data raises another factor for consideration: the role that instructional faculty play by including librarians and the library in the design and delivery of assignments. By integrating the library into the course, faculty connect their students with the

library as an academic support service. The results of this study suggest that faculty engagement with an instruction librarian and integration of library instruction into a course also is correlated with student retention results.

The study results show a downward trend in the number of students reached over the period discussed in this paper. One possible reason might be the rise in online instruction, which did not begin in earnest at GVSU until 2016. Librarians are much more likely to create asynchronous learning objects for online courses, which are more challenging to track as teaching faculty may reuse and share them without notifying the librarian. It is also possible that the downward trend is connected to shortages and changes in personnel that have affected the Libraries' capacity for instruction and outreach. There have also been some environmental factors, such as large-scale class cancellations due to inclement weather and the COVID-19 pandemic. However, it is difficult to pinpoint the definitive cause; it is more likely the result of several coinciding factors.

The possibility that engagement with the library is a high-impact educational practice has compelling implications.<sup>19</sup> One of the thirty assessment questions that University Libraries and IA explored asks, "Is there a relationship between *faculty* who work with a librarian on library instruction and the retention of *students who have that faculty member*?"<sup>20</sup> The answer, as shared previously by O'Kelly, is yes.<sup>21</sup> Not only is there a significant correlation between retention and having a faculty member who asked a librarian to teach an information literacy session, but also the odds ratio is similar. Additional research is needed to further test the null hypothesis that there is no relationship between faculty inviting library instruction and student retention; if proven false, it could indicate that encouraging student engagement with the library via library instruction is a faculty best practice and has a place alongside other high-impact practices such as undergraduate research, writing-intensive courses, and first-year experiences that were identified by George Kuh.<sup>22</sup>

### *Impact on Instructional Practices*

The results of the study have had a direct impact on the instructional practices at GVSU Libraries and demonstrate how analytics studies may inform day-to-day instructional practices while also communicating impact data to Libraries and campus administrators.

The analysis reports generated by IA show the number of unique students reached and the number of instruction sessions held, making it possible to track larger trends of engagement with library instruction across the university over time (Table 2).

<b>Year</b>	<b>Unique Students Reached</b>	<b>Instruction Sessions</b>
2015–2016	8,799	637
2016–2017	9,558	755
2017–2018	8,556	691
2018–2019	7,979	629
2019–2020*	7,930	429

\*COVID-19 ended library instruction mid-March 2020 through the end of the academic year, lowering the number of library instruction sessions for the year.



The analysis reports generated by IA identified academic programs that librarians were not reaching. This information triggered an opportunity for targeted outreach to those programs. The longitudinal tracking of overall instruction sessions allows the Libraries to view trends in library involvement with courses and programs over time. These data make the identification of pattern changes possible early and allow librarians the opportunity to ask questions about what might be changing in order to support a department's evolving needs.

<b>Number of Students Reached by Class</b>	<b>Total Students</b>	<b>Number of Students in Library Courses</b>	<b>% of Students in Library Courses</b>
Freshmen	4,904	2,968	61%
Sophomore	6,412	1,768	28%
Junior	7,595	1,613	21%
Senior	10,441	2,282	22%
Masters	3,613	723	20%
Doctoral	290	62	21%

The GVSU Libraries are also better equipped to strategically integrate GVSU's information literacy outcomes throughout the instruction program. One area of opportunity stems from the results in the data on the number of students reached at each grade level (Table 3). This data informs current strategizing on programmatic approaches to information literacy integration across all disciplines. The numbers of current academic-year-level integration support work in identifying information literacy-based learning objectives connected to year in school with the ultimate goal to more systematically incorporate higher-level information literacy competencies into upper-level courses scaffolded by consistent integration of introductory concepts in earlier years.

The data collected by the Libraries also measures the use of library instruction labs. Monitoring the percentage of instruction sessions held in library spaces provides evidence-based rationales for the use of spaces and in making funding requests to refurbish library instruction spaces and add new or additional resources in these spaces, as well as showing the need to retain those spaces specifically for library instruction purposes.

The data gathered at the Libraries-wide level and the strong relationship with IA also allows for individual instruction librarians to more easily work with IA to dig deeper into their own disciplinary instruction. By performing a more granular analysis of the data for their own disciplines than that analyzed at the programmatic level, instruction librarians have gained further insights into their individual instruction methods and subsequent impact.

### ***Limitations***

There are limitations to note with this study. It measures correlation, and while confounding factors such as intrinsic motivation, ACT score, high school GPA, socioeconomic status, and first-generation status were controlled via study design and analysis, this does not prove that library instruction causes an increase in student reenrollment the following fall. The large

population, length, and replicability of the study indicate something interesting is happening; further research is needed to determine what is causing that interesting relationship.

The particular institutional context may also limit the generalizability of the results to other institutions and, while the process was designed to be reproducible, the results themselves are not generalizable to all academic libraries.

Because we gathered the data through self-reported instructional data, this study is limited by human error at the data collection stage. For instance, if a librarian incorrectly enters a course code, this could have a minor impact on the final results. Although errors are mitigated by IA's cross-referencing of the course information with the lead instructor name, some entry errors may still persist into analysis.

As mentioned earlier, the librarians intentionally do not take attendance during the instruction sessions, so the attendance is estimated based on the overall enrollment in the course, not the actual students who attended the instruction session. Although the Institutional Analysis department is confident that typical absences for a course fall within the margin of error for analyses like these, it does mean that students who were absent on the day the library session occurred may have been included in the retention analysis.

Similar to the attendance limitation, there was no non-intrusive way within the scope of this study to determine how attentive or engaged students were during library instruction sessions. It is possible that students who were in attendance at the library instruction session were otherwise distracted or unengaged; they physically may have been present but were not learning the material, which therefore limits any implication that there is a relationship between actual internalized learning of library instruction content and subsequent reenrollment.

The study does not account for online instruction. Online instruction did not begin in earnest at GVSU until after 2016, with complete online instruction during the COVID-19 closure between mid-March and August 2020. Online instruction, therefore, did not represent a large enough sample to analyze. Online and hybrid instruction is a growing area at GVSU, with ongoing strategic emphasis on growing the Libraries' capacity and expertise in online instruction, so its impact on student learning could be explored in greater detail in future studies but is necessarily absent from this study.

There are also limitations in how we define retention; in this case, retention is not persistence to graduation, but rather a return to the university the following fall semester. This study does not account for graduation rates. The study also notes that decisions to withdraw from a specific academic institution are often complicated and multifaceted, as there are many reasons or combinations of reasons that students choose to return or withdraw, independent of library engagement.

## Conclusion

The retention analytics methodology outlined by this eight-year study provides a reproducible path for other institutions to partner with their local institutional analysis colleagues to create a study that explores the relationship between library instruction and student reenrollment while also protecting individual students' data. By focusing on students in enrolled classes with mandatory in-class instruction sessions instead of voluntary workshop or circulation data, the methodology also reduces the likelihood that students' intrinsic motivation will indicate an over-reporting of impact in the findings of libraries and student success.

This study demonstrates a positive correlation between library-led instruction and student reenrollment the following fall at GVSU, along with a reproducible and reliable method for measuring that correlation. Although the results do not suggest a causal relationship, the large population studied and the consistency of the results over time warrant deeper exploration in order to further define which specific factors are influencing the results. By replicating the process at different institutions and different instruction models (such as online), and by designing future research studies that dive more deeply into the potential significance of faculty engagement with libraries as a high-impact practice that positively influences student success, more details may emerge that help explain the phenomenon of higher fall-to-fall retention and library instruction.

## Appendix A

### Research Questions

1. How many students did we reach in total? And per cent?
2. How many students did we reach in each of the programs (AAA, HTM, HST, etc.)? And per cent?
3. How many freshmen, sophomores, juniors, seniors, graduate, etc. in total? And per cent?
4. How many freshmen, sophomores, juniors, seniors, graduate, etc. in each of the programs? And per cent in each program?
5. Of the students who saw a librarian in 2020–2021, what percentage of them re-enrolled for F21, compared to students who did not see a librarian? (overall and by grade FR SO JR SR GM GD)
6. How many programs did librarians reach?
7. How many programs are not listed here?
8. Which programs are not listed here?
9. How many unique classes did librarians visit? (Some librarians visit a class multiple times, and will have same course code/number/section within one semester.)
10. How many students saw a librarian more than once?
11. What are the characteristics of those students who saw a librarian more than once (classes, year in school, major)?
12. What is the highest number of times that a student saw a librarian?
13. Is there an intensity effect on reenrollment or GPA of students who saw a librarian multiple times?
14. What number and percentage of freshman take WRT 098? And WRT 150?
15. What number and percentage of those freshman did we reach in the WRT 098 classes? And the WRT 150 classes?
16. What percentage of WRT 150 students who saw a librarian re-enrolled for F15, compared to those students in WRT 150 who did not see a librarian?
17. What percentage of first-generation freshmen take WRT 098 and WRT 150? How many did we reach?
18. What percentage of library sessions were in the library (“location” field)? What percent online?
19. How many hours of instruction in total (“duration”) did librarians provide in 2020–2021?
20. Can you create a line graph showing the number of instruction sessions per day over the course of the academic year?
21. How many sessions did each librarian teach, sorted by librarian?
22. How many sessions did each librarian teach in each program, sorted by librarian?
23. How many students did each librarian teach, sorted by librarian?
24. Can you create a bell curve of the distribution of the number of sessions each librarian teaches?
25. What are the majors of the students who enrolled in WRT150? What are the majors of the students who enrolled in WRT150 AND saw a librarian?
26. What are the top ten courses at the university that reach the most freshmen?

## **Correlation Between Library Instruction and Student Retention 97**

27. P value for the library instruction reenrollment correlation AY2020–2021 (significance level .05)
28. Total number of faculty at the university each of those three years
29. Retention of (all students who were taught by faculty who had a librarian come into class) compared to retention of (all students who did not have faculty who had a librarian come to class) for AY 2020–2021
30. Variables to control: ACT score; high school GPA; low income; first generation

## Appendix B

### SAS Script

The IA analyst used the following script in SAS Enterprise Guide 7.1 to analyze the data:

Chi-square with odds ratio:

```
proc freq;
table library*retent/chisq or;
run;
```

*Note: Library is coded 1 for "student who was in library instruction section" and 0 for "student who was not in library instruction section."*

*Note: Retention is coded 1 for "student was retained the following fall semester" and 0 for "not retained in the follow fall semester."*

Generalized linear model:

```
proc glm data=retention;
class library firstgen pellever;
model retent=library actecomp hsgpa1 firstgen pellever/solution;
run;
```

actecomp – Comprehensive ACT score

hsgpa1 – HS GPA

firstgen – First generation student (1=Yes 0=No)

pellever – Student is pell-eligible (1=Yes 0=No)

### Notes

1. Preliminary, partial results of this ongoing study have been presented and published in other venues. This is the first complete peer-reviewed publication of the methods and data for all years 2012–2020.

2. George D. Kuh and Robert M. Gonyea, "The Role of the Academic Library in Promoting Student Engagement in Learning," *College & Research Libraries* 64, no. 4 (July 2003): 256–82. <https://doi.org/10.5860/crl.64.4.256>.

3. Megan Oakleaf, *Value of Academic Libraries: A Comprehensive Research Review and Report*, Chicago: Association of College and Research Libraries, 2010. [http://www.ala.org/acrl/sites/ala.org/acrl/files/content/issues/value\\_val\\_report.pdf](http://www.ala.org/acrl/sites/ala.org/acrl/files/content/issues/value_val_report.pdf).

4. Ibid.

5. Ibid.

6. Melissa A. Hubbard and Amber T. Loos, "Academic Library Participation in Recruitment and Retention Initiatives," *Reference Services Review* 41, no. 2 (June 2013): 157–81. <https://doi.org/10.1108/00907321311326183>.

7. Shun Han Rebekah Wong and T.D. Webb, "Uncovering Meaningful Correlation between Student Academic Performance and Library Material Usage," *College & Research Libraries* 72, no. 4 (July 2011): 361–70. <https://doi.org/10.5860/crl-129>; Krista M. Soria, Jan Fransen, and Shane Nackerud, "Stacks, Serials, Search Engines, and Students' Success: First-year Undergraduate Students' Library Use, Academic Achievement, and Retention," *Journal of Academic Librarianship* 40, no. 1 (January 2014): 84–91. <https://doi.org/10.1016/j.acalib.2013.12.002>; DeeAnn Allison, "Measuring the Academic Impact of Libraries," *Portal: Libraries and the Academy* 15, no. 1 (January 2015): 29–40. <https://doi.org/10.1353/pla.2015.0001>; John K. Stemmer and David M. Mahan, "Investigating the Relationship of Library Usage to Student Outcomes," *College & Research Libraries* 77, no. 3 (May 2016): 359–75. <https://doi.org/10.5860/crl.77.3.359>; Angie Thorpe, Ria Lukes, Diane J. Bever, and Yan He, "The Impact of the Academic Library on Student Success: Connecting the Dots," *Portal: Libraries and the Academy* 16, no. 2 (April 2016): 373–92. <https://doi.org/10.1353/pla.2016.0027>.

8. Gaby Haddow and Jayanthi Joseph, "Loans, Logins, and Lasting the Course: Academic Library Use and Student Retention," *Australian Academic & Research Libraries* 41, no. 4 (December 2010): 233–44. <https://doi.org/10.800048623.2010.10721478>; Gaby Haddow, "Academic Library Use and Student Retention: A Quantitative Analysis," *Library & Information Science Research* 35, no. 2 (April 2013): 127–36. <https://doi.org/10.1016/j.lisr.2012.12.002>; Adam Murrar, Ashley Ireland, and Jana Hackathorn, "The Value of Academic Libraries: Library Services as a Predictor of Student Retention," *College & Research Libraries* 77, no. 5 (September 2016): 631–42. <https://doi.org/10.5860/crl.77.5.631>.

9. Gregory A. Crawford, "The Academic Library and Student Retention and Graduation: An Exploratory Study," *Portal: Libraries and the Academy* 15, no. 1 (January 2015): 41–57. <https://doi.org/10.1353/pla.2015.0003>.
10. Mark Emmons and Frances C. Wilkinson, "The Academic Library Impact on Student Persistence," *College and Research Libraries* 72, no. 2 (March 2011): 128–49. <https://doi.org/10.5860/crl-74r1>.
11. Elizabeth M. Mezick, "Return on Investment: Libraries and Student Retention," *Journal of Academic Librarianship* 33, no. 5 (September 2007): 561–66. <https://doi.org/10.1016/j.acalib.2007.05.002>.
12. Krista M. Soria, Kate Peterson, Jan Fransen, and Shane Nackerud, "The Impact of Academic Library Resources on First-Year Students' Learning Outcomes," *Research Library Issues* 290 (2017): 5–20. <https://doi.org/10.29242/rli.290>.
13. Graham Stone and Bryony Ramsden, "Library Impact Data Project: Looking for the Link between Library Usage and Student Attainment," *College & Research Libraries* 74, no. 6 (November 2013): 546–59. <https://doi.org/10.5860/crl12-406>.
14. Silas M. Oliveira, "The Academic Library's Role in Student Retention: A Review of the Literature," *Library Review* 66, no. 4/5 (July 2017): 310–29. <https://doi.org/10.1108/LR-12-2016-0102>.
15. M. Brooke Robertshaw and Andrew Asher, "Unethical Numbers? A Meta-analysis of Library Learning Analytics Studies," *Library Trends* 68, no. 1 (Summer 2019): 76–101. <https://doi.org/10.1353/lib.2019.0031>.
16. Megan Oakleaf, Anthony Whyte, Emily Lynema, and Malcolm Brown, "Academic Libraries & Institutional Learning Analytics: One Path to Integration," *The Journal of Academic Librarianship* 43, no. 5 (September 2017): 454–61. <https://doi.org/10.1016/j.acalib.2017.08.008>.
17. Sarah Prindle and Amber Loos, "Information Ethics and Academic Libraries: Data Privacy in the Era of Big Data," *Journal of Information Ethics* 26, no. 2 (Fall 2017): 22–33; Kyle M. L. Jones and Dorothea Salo, "Learning Analytics and the Academic Library: Professional Ethics Commitments at a Crossroads," *College & Research Libraries* 74, no. 3 (April 2018): 304–23. <https://doi.org/10.5860/crl.79.3.304>; Kyle M. L. Jones, "'Just Because You Can Doesn't Mean You Should': Practitioner Perceptions of Learning Analytics Ethics," *Portal: Libraries and the Academy* 19, no. 3 (July 2019): 407–28. <https://doi.org/10.1353/pla.2019.0025>; Robertshaw and Asher, "Unethical Numbers?," 76–101; Alan Rubel and Kyle M.L. Jones, "Student Privacy in Learning Analytics: An Information Ethics Perspective," *The Information Society* 32, no. 2 (February 2016): 143–59. <https://doi.org/10.1080/01972243.2016.1130502>.
18. Kristin A. Briney, "Data Management Practices in Academic Library Learning Analytics: A Critical Review," *Journal of Librarianship and Scholarly Communication* 7, no. 1 (February 2019): 1–39. <https://doi.org/10.7710/2162-3309.2268>.
19. Mary O'Kelly, "Academic Libraries and Student Retention: The Implications for Higher Education," in *Proceedings of the 2016 Library Assessment Conference: Building Effective, Sustainable, Practical Assessment*, ed. Sue Baughman, Steve Hiller, Katie Monroe & Angela Pappalardo (Washington, DC: Association of Research Libraries, 2017), 485–90. <https://www.libraryassessment.org/wp-content/uploads/bm~doc/proceedings-2016.pdf>
20. This was presented to the analyst as, "What is the retention of (all students who were taught by faculty who had a librarian come into class) compared to retention of (all students who did not have faculty who had a librarian come to class) for academic year XXXX-XXXX?"
21. O'Kelly, "Academic Libraries and Student Retention," 485–90.
22. George D. Kuh, *High-Impact Educational Practices: What They Are, Who Has Access to Them, and Why They Matter*, Washington, DC: Association of American Colleges and Universities, 2008.