Data Literacy Practices of Students Conducting Undergraduate Research

Theresa Burress

Undergraduate research is considered to be a high-impact practice; however, research into the data literacy of students conducting undergraduate research is lacking. In addition, institutionwide assessments of data practices are challenging because of varied disciplinary approaches to data. This study investigates the data practices of undergraduate students who submitted research posters to a student research symposium in 2019 and found that students engage in a variety of data practices during undergraduate research regardless of their research method and approach for obtaining and working with data. This study identifies potential areas for data-related library instruction in support of undergraduate research.

Undergraduate research is widely considered to be a high-impact practice that increases the interest and engagement of students.[1] As such, many colleges and universities are working to expand access to undergraduate research opportunities. When students first engage in original research, they cross the threshold from being knowledge consumers to becoming knowledge creators. Information and data literacy are integral to the research experience, and students often disseminate their findings at institutionwide symposia by presenting research posters with data visualizations that represent an array of data practices.

Data literacy is a multifaceted set of skills that involve understanding and using data effectively.[2] Prado and Marzal's broad framework of data-related competencies situates data literacy in the context of information literacy instruction in all types of libraries. Many data literacy competencies require knowledge practices akin to those outlined in the Association of College and Research Libraries' *Framework for Information Literacy*.[3] Thus, integrating information and data literacy throughout the curriculum provides essential preparation for undergraduate students who are increasingly likely to conduct original research.

While broad assessments of undergraduate research have been conducted across higher education,[4] and discipline-specific assessments have been done within the classroom,[5] no empirical studies have explored the data practices of students who have disseminated undergraduate research at campuswide symposia. This may be because institutionwide assessments of undergraduate research are challenging, in part due to the varied disciplinary approaches to data and research. Faculty may recommend or require that students use specific research methodologies, datasets, or data analysis tools. The structure and extent of the project may

---

also vary depending upon whether the research is conducted as part of a course, directed individual study, a thesis, or part of a research assistantship.

This study uses a data literacy framework adapted for a college campus[6] and further refined for undergraduate curricula[7] to investigate the data literacy of students who submitted research posters to a campuswide research symposium in 2019. The findings offer insights into the data practices of students who are conducting undergraduate research and will inform the development of academic programming and tools to improve and assess the data literacy of undergraduate researchers.

Literature Review

Undergraduate research is a practice through which undergraduate students conduct a research inquiry or investigation that results in "an original intellectual or creative contribution" to a discipline.[8] While undergraduate research programs have operated at colleges and universities for decades, undergraduate research was identified as one of 10 high-impact practices in 2008 by George Kuh, who led research that identified specific activities (such as undergraduate research, service learning, and internships) as being particularly beneficial in improving the retention and success of undergraduate students from many different backgrounds.[9]

Undergraduate research has been assessed using a variety of approaches at different scales. National surveys have been used to measure student-reported gains resulting from their participation in undergraduate research, and the results indicate that students who report higher benefits of the experience tend to be more likely to pursue subsequent research opportunities via advanced degree programs.[10] Other studies have used the narrow scope of a single course, such as a study by Bracher, Cantrell, and Wilkie, that used a poster presentation assignment to assess learning outcomes such as critical inquiry and dissemination of findings.[11] Fewer studies have taken an institutionwide approach. One study with a universitywide scope investigated perceptions of undergraduate students involved in research across a variety of settings (such as course projects, honor theses, and independent study) and found that the undergraduate research experience significantly improved student perceptions of their understanding of the research process, confidence in taking initiative, and presentation skills.[12] Another universitywide study surveyed freshman and sophomore students with early research experiences in STEM and non-STEM disciplines and found few major differences in the learning gains reported by STEM and non-STEM students.[13] As student research becomes more prevalent in the undergraduate curriculum, Kezar and Holcomb assert that more direct measures to assess student learning at the institutional level are needed.[14]

Undergraduate research programs represent a key area of engagement by academic librarians, who aim to advance institutional priorities by building deep collaborations[15] on issues ranging from research to teaching and learning.[16] The Council on Undergraduate Research recognizes the need for adequate library resources and suggests that "support for information-literacy training and development of research skills should be built into the curriculum" to ensure the success of undergraduate research programs;[17] however, the role of librarians as engaged collaborators in instruction, programming, and assessment is not explicitly defined. The *Framework for Information Literacy*[18] articulates the central role of information literacy throughout the research process, and research done by Hensley, Shreeves, and Davis-Kahl shows that libraries at a range of higher education organizations support formal undergraduate research programs with dedicated space, collections, and instruction-related activities.[19]

Research related to the data literacy of students who are conducting research in an undergraduate setting is lacking. Hensley surveyed librarians to better understand the role of library instruction in formal undergraduate research programs and found that nearly 93 percent of respondents provide specialized instruction.[20] She found that the most frequent information literacy topic taught is traditional database searching and techniques (13%); 5 percent of respondents indicated that they taught searching for statistical information, but all other data-related topics (such as numeric and geospatial data, data visualization techniques, and developing data management plans) were mentioned by 1 percent or fewer respondents. Hensley includes data management among several information literacy topics that under-graduate researchers "may be ready to delve into, albeit at a beginner's level."

Academic librarians working in the area of data literacy often focus on aspects of research data management, as in the Data Information Literacy Project.[21] In the initial needs assessment for that project, Carlson et al. proposed a set of competencies around the term "data informa-tion literacy," or DIL, which merges the idea of "researcher-as-producer" of data products with the idea of "researcher-as-consumer" of data products.[22] These DIL competencies were created for the specialized context of training new researchers and graduate students in e-research, primarily within the sciences. In looking toward future work, Carlson wrote that introducing DIL to undergraduate students could be useful but acknowledged that tailoring such programs for undergraduate settings would be challenging because most undergraduates do not produce datasets. However, he suggested that undergraduate research programs could serve as points of entry for DIL and proposed a number of opportunities for future research, in particular investigating the "contextual aspects of data skills" and exploring "students' relationships to the data they are generating or working with."[23]

As the need for data literacy instruction in the undergraduate curriculum increases, as documented in a case study by Battista, Boss, and McCartin,[24] so has its relevance and interest to academic librarians. In the final chapter of her 2021 monograph, Julia Bauder addresses in depth how the principles of the *Framework for Information Literacy* can further inform data literacy pedagogy.[25] Bauder states, "In many ways, thinking critically about data involves the same questions as thinking critically about texts," and goes on to map several key questions about data to the six frames that comprise the *Framework*. Burress, Mann, and Neville first investigated the data literacy of undergraduate students within the structure of a faculty learning community, adapting and customizing data literacy competencies for a midsized college campus that primarily serves undergraduate students.[26] Rather than using DIL for this project, they used a definition of data literacy that situates the concept along a continuum with information literacy. Prado and Marzal defined data literacy as "the component of information literacy that enables individuals to access, interpret, critically assess, manage, handle, and ethically use data."[27] The associated competencies are meant to be flexible and adaptable by librarians who wish to integrate data literacy into their information literacy instruction. While Prado and Marzal acknowledge the interdependence between data and statistical literacy, they propose that data literacy is the "umbrella concept" that includes statistical literacy rather than the reverse.[28] Burress, Mann, Montgomery, and Walton built on this work with a study that investigated data literacy teaching in the undergraduate classroom at two institutions. They found that faculty across disciplines largely agreed that most data literacy competencies are relevant in the undergraduate classroom.[29]

The current investigation complements and builds on previous studies by examining the

data practices and perceptions of students who have conducted undergraduate research in a variety of disciplines.

Methodology

This study investigates data practices of students who presented research posters at a campus-wide symposium held at a midsized branch campus of a public research university in 2019. The author's research objectives were to identify the range of data practices that students use while conducting undergraduate research and determine to what extent these practices align with student perceptions and faculty priorities. This study is guided by the following hypotheses:

H1. Students who design an experiment and/or collect original data engage in more data practices than students who use external or compiled datasets.

H2. Students who use quantitative data analysis techniques engage in more data practices than students who use other techniques to analyze data.

H3. The use of data practices, such as obtaining data, evaluating data, analyzing data, creating visualizations, citing data, cleaning/converting data, and creating metadata, are documented on undergraduate research posters.

H4. Selected data literacy competencies deemed relevant for the undergraduate curriculum by faculty in an earlier study, including obtaining data, evaluating data, analyzing data, creating visualizations, citing data, cleaning/converting data, are also relevant from an undergraduate researcher perspective.

Participants and Procedures

At the time of the data collection, the university was a separately accredited master's-level institution serving approximately 4,500 students with 28 undergraduate and 18 graduate programs in Colleges of Arts and Sciences, Business, and Education. For 15 years, the campus Annual Student Research Symposium has been organized by the Office of Research to provide students with a forum to present their original research in a "public demonstration of competence."[30] The author worked with a librarian colleague and Office of Research staff to design the survey and coordinate data collection, including survey responses, symposium application, and research poster files. The research methodology was reviewed and deemed exempt by the university's Institutional Review Board.

In all, 154 research posters were submitted to the 2019 University Student Research Symposium. Research posters represented individual and coauthored projects conducted as part of courses, directed individual study, honors theses, and campus research lab work. Some research posters included faculty coauthors. The symposium is open to all students at this campus; however, the College of Education organizes a separate symposium to accommodate the schedules of working teachers. Therefore, this dataset represents research undertaken by students in the College of Arts and Sciences and the College of Business.

Research participants were recruited from the overall group of symposium participants. Most respondents completed the survey on site during the event, and the remaining respondents completed the survey online upon receiving an electronic survey link distributed via email during the week after the symposium. Survey data was collected from 74 symposium participants using Qualtrics survey software; survey responses were matched to symposium application data and electronic poster files via applicant names and poster titles. Survey responses that were not able to be matched to a poster were discarded. In addition, 10 graduate student

survey responses and associated poster submissions were removed from the dataset to focus this research strictly on the undergraduate experience. The total number of participants for this study included 63 undergraduate students and 58 corresponding digital research posters.

Measurement Instruments

The survey was developed collaboratively by the author, another librarian, and Office of Research partners. The first section of the survey asked for information regarding participants' major area of study, motivations, and preparation for their research project, including research methods coursework and previous involvement in other high-impact practices. The second section of the survey asked about the participants' engagement with specific data practices. The third section of the survey asked about the students' perception of the challenges they faced during the project, how they may have used library resources, and their perception of their improvement with regard to data-related skills.

This paper focuses primarily on the survey data taken from the second section of the survey regarding engagement with specific data practices. Column 1 of table 1 outlines selected data literacy competencies that are likely to be used during undergraduate research. Column 2 includes the exact language used in the survey to ask whether respondents engaged in specific data practices while working on their research.

Data Collection Analysis

The author collected the poster data by closely reviewing the electronic poster files and documenting proxy evidence of the following data practices: characteristics of data sources (such as collection of new/original data, finding and using an external dataset, compilation of a dataset from multiple sources), evidence of quantitative data analysis as defined by Leavy,[31] number and type of data visualizations, and whether the data source was cited appropriately if published data were used. The poster review also explored whether proxy evidence of dataset evaluation, dataset cleaning/converting, and creation of metadata could be discerned. The survey results were then compared with the proxy evidence to determine whether the student reports agreed with the evidence collected from the posters.

For some variables, evidence from the research posters was explicitly documented. For others, the author proposes poster characteristics that may be used as a proxy. Column 3 of table 1 describes the direct or proxy characteristic used to identify the presence or absence of each data practice.

TABLE 1 Data Literacy Competencies and Related Practices: Reported & Proxy Evidence		
Selected Competencies[32]	**Survey asked:** "While working on my research, I: (check all that apply)"	**Proxy Evidence of Data Practices**
Find, select, access, or create datasets to test a hypothesis or answer a research question	Collected new data	Poster reports new data
	Found and used an existing dataset	Poster reports the use of an existing data source
	Compiled a dataset from multiple existing data sources	Poster describes compilation of multiple existing data sources
	None (i.e., used no data)	None (i.e., used no data)

TABLE 1 (CONTINUED) Data Literacy Competencies and Related Practices: Reported & Proxy Evidence		
Selected Competencies[32]	Survey asked: "While working on my research, I: (check all that apply)"	Proxy Evidence of Data Practices
Interpret and critically evaluate data & their sources	Evaluated the quality of the dataset	Poster displayed data from one or more external sources (i.e., existing, compiled data source)
Analyze data	Analyzed my data	Text or visualizations describe quantitative or qualitative data analysis method
Ethically collect/use/cite data	Cited and/or obtained permission to use my data sources	If poster describes existing or compiled data sources, source(s) are cited
Communicate data effectively to different audiences in part by using visualizations	Created graphs, charts, illustrations, figures, etc.	At least one or more original data visualization(s) are used
Clean/process/convert data	Cleaned and/or converted my data to a different format	Poster describes the use of a data analysis tool or software (e.g., Excel, JMP, etc.)
Create metadata to meet data publication requirements	Created codes or tags to describe data (i.e., metadata)	Poster describes the creation of metadata, codes, or tags for data

Study Limitations

Because the author was the only researcher involved in devising the codes and generating the resulting data, it should be acknowledged that other researchers may have made different choices with regard to the coding scheme as well as the resulting categorization of data. It was challenging to identify appropriate proxy evidence that demonstrated each data competency and associated practice. For example, in classifying data source types that were displayed on each poster, the broadest possible conception of the term "data" was used to accommodate unstructured, or textual, data[33] as well as structured, numerical data. Historical and literary primary source material was classified as a compiled dataset. While students and practicing researchers in many disciplines often think only of numerical data as "data," the broader conception of data used in this investigation is meant to better accommodate the various types of qualitative and unstructured data that researchers in the social sciences and humanities may collect and analyze.

In the Interpret and Critically Evaluate Data category, the attempt to identify feasible proxy evidence was unsuccessful. The presumption that students evaluated datasets if they used external or compiled datasets as proposed in table 1 was found to be inadequate. Close review of the posters showed that all students submitting posters for one specific course used the same external, unpublished dataset provided by their instructor; thus, these students used an external dataset but did not make an evaluative judgment about the data source.

Finally, the attempt to identify feasible proxy evidence for the Create Metadata category was unsuccessful. The research posters did not contain explicit documentation of this data practice. Teaching faculty often use a scaffolding approach to major research assignments,

which may require student reports about data source evaluation, selection, and the creation of metadata. Coordination with faculty advisors to collect this type of supplemental data has the potential to improve this study design.

Findings

Characteristics of Participants and Posters

The data analyzed for this study includes: (1) 63 survey responses from undergraduate students, and (2) 58 corresponding digital poster files with associated application data. All 63 research participants were undergraduate students over the age of 18. Most students (55; 87.3%) were first-time presenters at the Student Research Symposium. However, more than a third of the first-time presenters (20; 36.3%) reported that they had previously conducted undergraduate research. Tables 2 and 3 provide characteristics of research participants and posters, respectively. As shown in Table 2, the expected year of graduation varied for this group of student researchers. Most students (54; 85.7%) reported completing at least one research-related course prior to their research project, on average completing two such courses.

Table 3 shows that a majority (45; 73.0%) of the posters were completed as part of a course requirement. More than half of the posters (36; 57.1%) had at least two authors. While the students' major field of study was aligned with the discipline of their research poster in most cases, a substantial minority of student researchers (13; 20.6%) conducted research in a discipline different from their reported major.

TABLE 2 Characteristics of Undergraduate Student Researchers		
Expected Year of Graduation	2019	40 (63.5%)
	2020	20 (31.7%)
	2021	3 (04.8%)
Research-related Course(s) Completed	At least one statistics course	54 (85.7%)
	At least one methods course	38 (60.3%)
	None	9 (14.3%)
Major Field of Study	Natural or Health Sciences	36 (57.1%)
	Social Sciences	26 (41.3%)
	Humanities	1 (01.6%)

TABLE 3 Research Poster Characteristics		
Research Project Setting	Course	45 (73.0%)
	Directed Individual Study	9 (14.3%)
	Honors Thesis	4 (06.3%)
	Lab Research Assistant	5 (07.9%)
Author(s)	One author	27 (44.4%)
	Two or more co-authors	36 (57.1%)
	At least one faculty co-author	9 (14.3%)
Discipline/Field of Project	Natural and/or Health Sciences	36 (57.1%)
	Social Sciences	26 (41.3%)
	Humanities	1 (01.6%)

Student Engagement in Data Practices
Survey Results

All but one participant (62; 98.4%) reported doing at least one data practice during the course of their research project; overall, students engaged in an average of four data practices, with two students reporting engagement with all nine data practices. Figure 1 shows the most frequently reported data practices, including obtaining data (59; 93.7%), analyzing data (48; 76.2%), and creating data visualizations (46; 73%). A majority of students also reported citing or obtaining permission to use data (35; 55.6%).

FIGURE 1
Student Survey Data Summarizing Percentages of Student Engagement in Data Practices
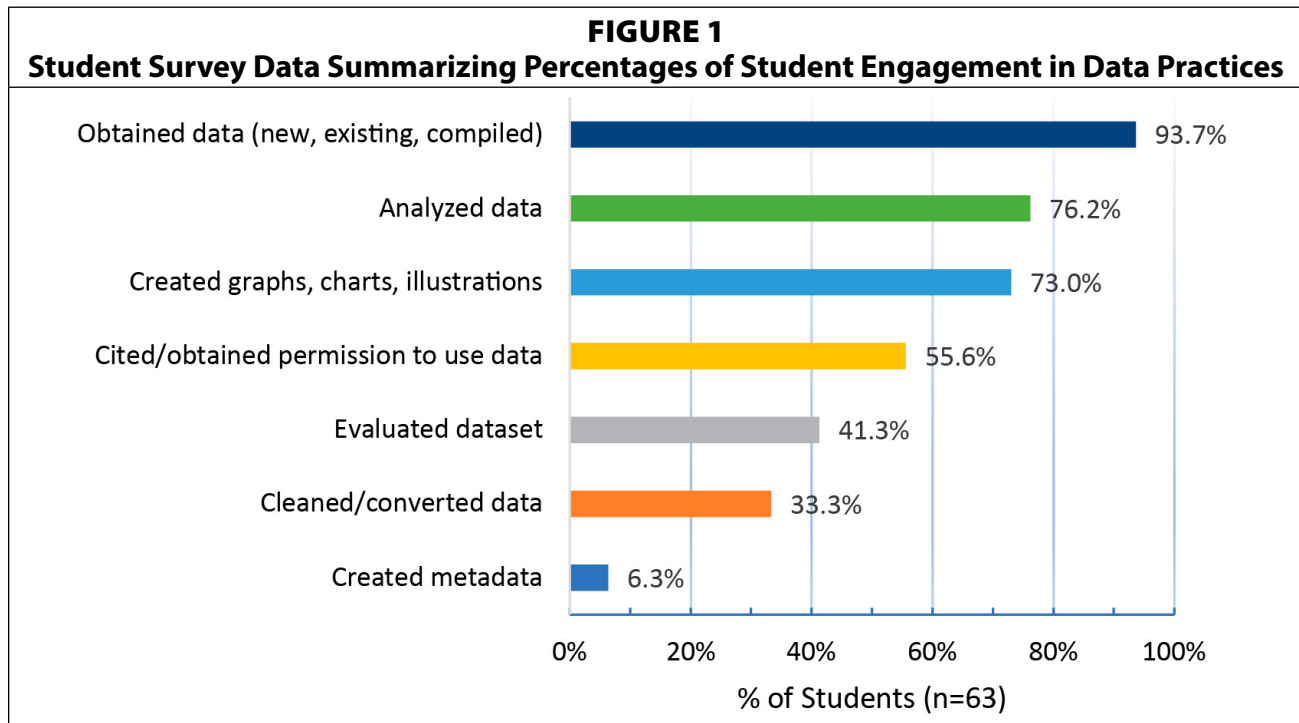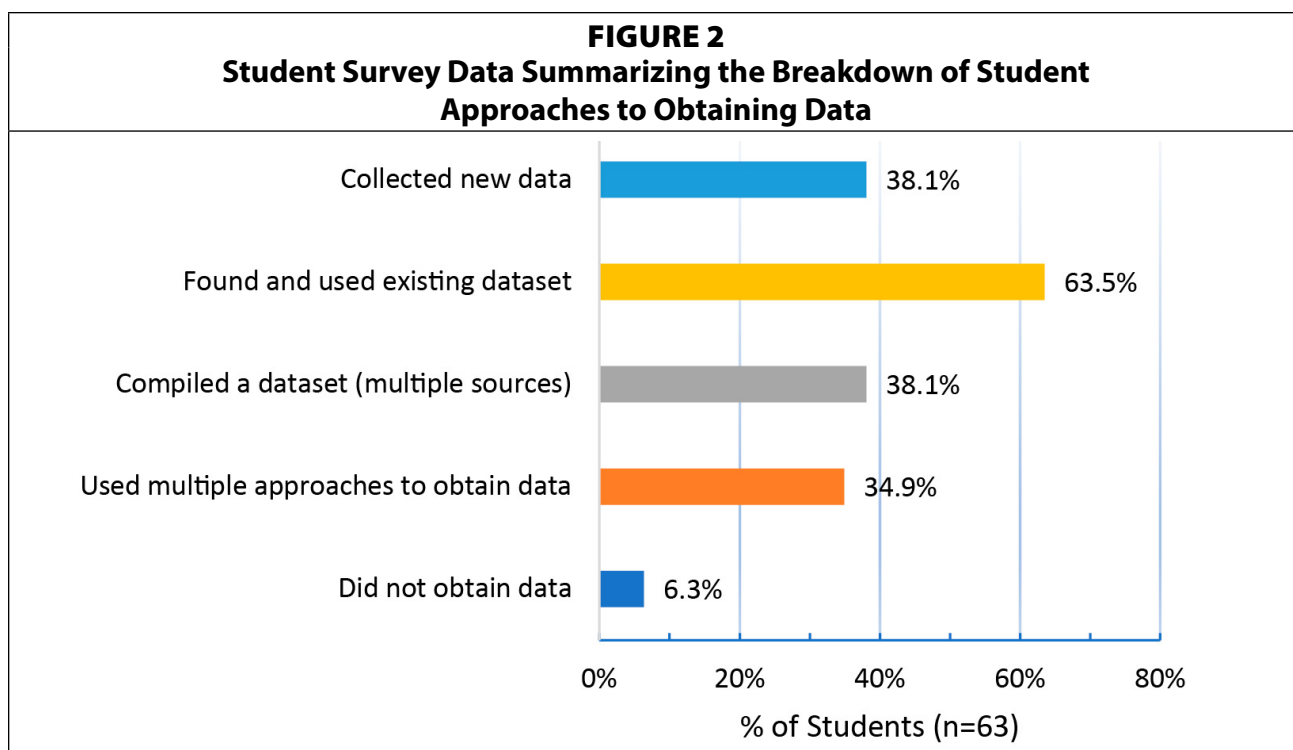

Most respondents (59; 93.7%) reported obtaining data using at least one approach (see figure 2). More than a third of students reported using at least two approaches to obtain data.
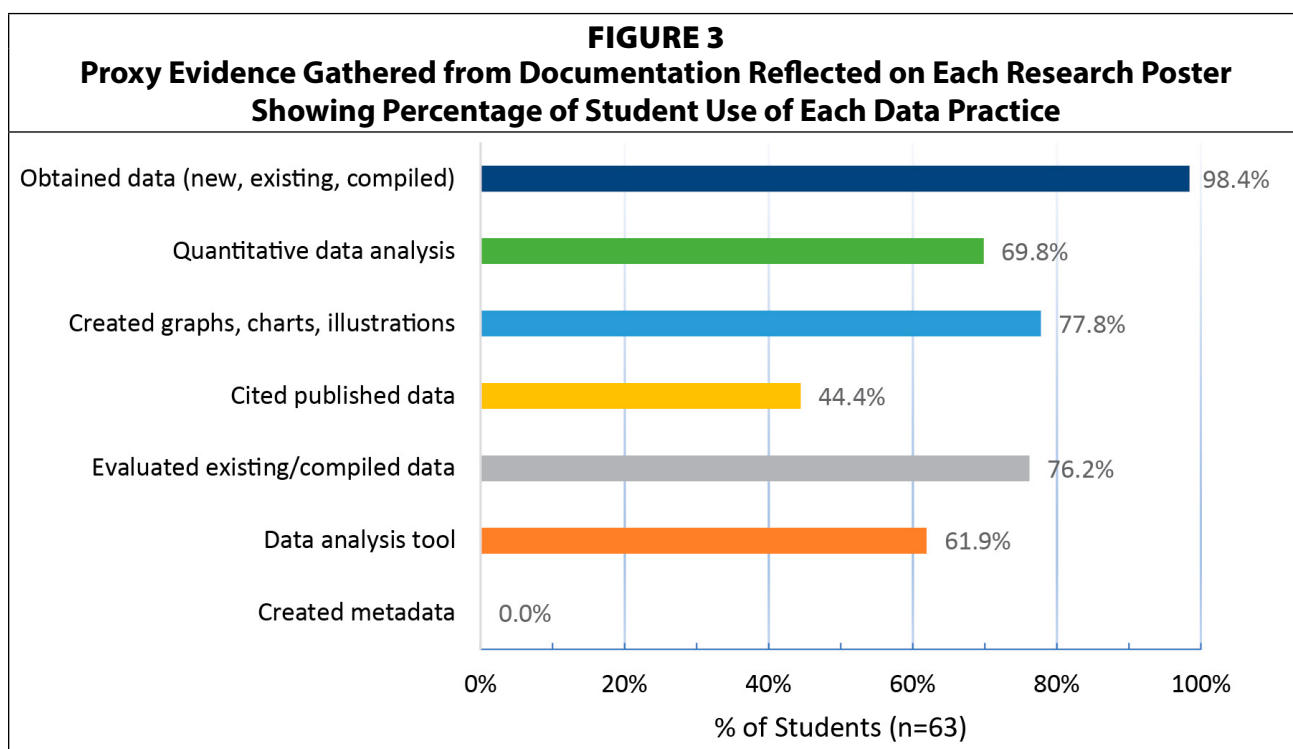
Content Analysis of Data Presented on Research Posters

Careful examination of the research posters indicated that students used a variety of research methods for their projects. The most frequently described methods included descriptive studies (30; 47.6%), historical or literacy analysis of primary source material (11; 17.5%), surveys (7; 11.1%), and computer models (6; 9.5%). Other students developed experimental methods, conducted qualitative or geospatial analysis, or built engineering prototypes. One poster described a review and synthesis of existing research.

Figure 3 shows the percentage frequency of proxy evidence for each data practice, the most prevalent being obtaining data (62; 98.4%) and creating data visualizations. Forty-nine students (77.8%) displayed at least one data visualization (such as table, graph, map, or timeline) on their posters; on average, four data visualizations were displayed on each poster. Proxy evidence showed that students used external data sources and therefore were likely to have evaluated those data sources (48; 76.2%). Proxy evidence of the use of quantitative data analysis techniques (44; 69.8%) was also prevalent. A majority of student posters (39; 61.9%)

FIGURE 2
Student Survey Data Summarizing the Breakdown of Student
Approaches to Obtaining Data

Collected new data — 38.1%
Found and used existing dataset — 63.5%
Compiled a dataset (multiple sources) — 38.1%
Used multiple approaches to obtain data — 34.9%
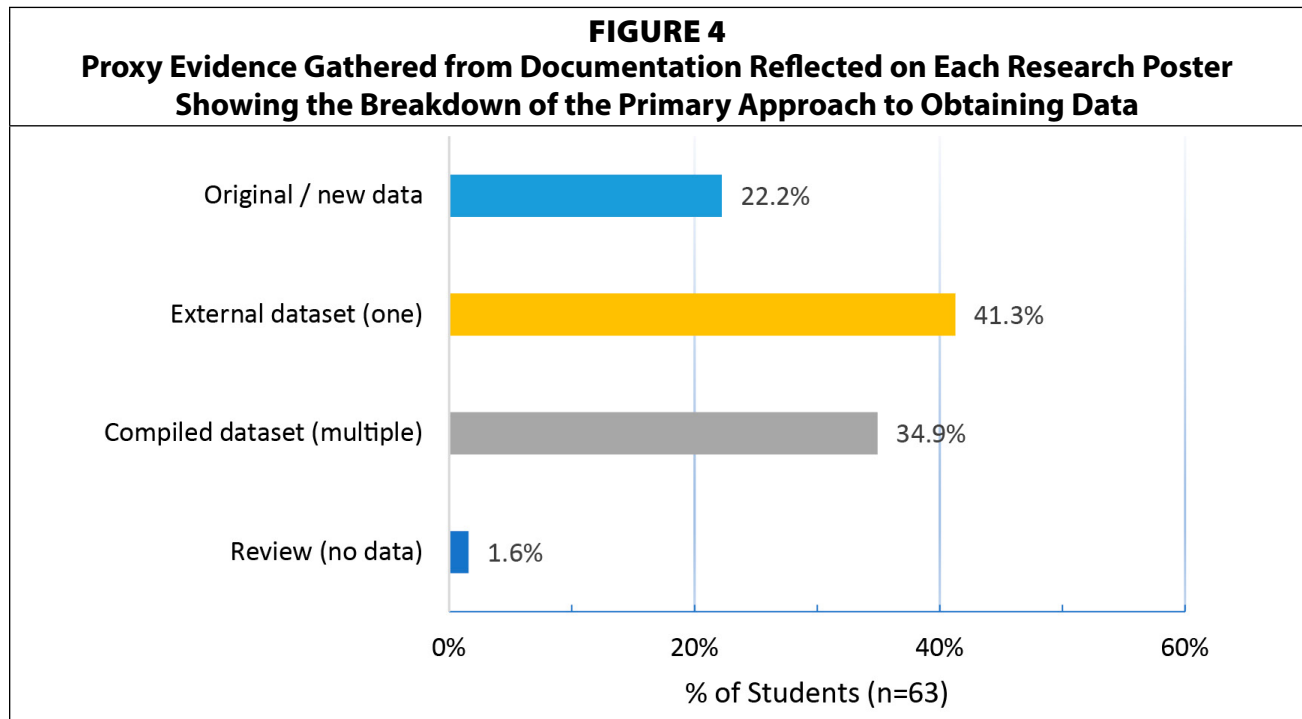Did not obtain data — 6.3%

% of Students (n=63)

documented the use of a data analysis tool such as Excel, JMP, photometer, or Arduino, which indicates that these students likely spent time cleaning or converting data to a specific format for the purpose of analysis. Percentage of proxy evidence reflecting appropriate data citation practices was calculated only on posters where external and/or compiled data sources were used (n = 36). Of this group, fewer than half of the students explicitly cited their data sources (16; 44%). Although a small percentage of students reported that they "created codes or tags

FIGURE 3
Proxy Evidence Gathered from Documentation Reflected on Each Research Poster
Showing Percentage of Student Use of Each Data Practice

Obtained data (new, existing, compiled) — 98.4%
Quantitative data analysis — 69.8%
Created graphs, charts, illustrations — 77.8%
Cited published data — 44.4%
Evaluated existing/compiled data — 76.2%
Data analysis tool — 61.9%
Created metadata — 0.0%

% of Students (n=63)

to describe data (i.e., metadata)," review of the posters did not present any corresponding evidence.

The data sources displayed on each poster were categorized as follows: 1) Original data (that is, collected and presented new data) (14; 22.2%); 2) External dataset from one source (26; 41.3%); 3) Compiled dataset from multiple external sources (22; 34.9%); or 4) Literature review only (that is, no data) (1; 1.6%) (see figure 4). Each poster was categorized with only one data source type; therefore, these percentages differ from the survey responses shown in figure 2, in which students were permitted to select multiple ways of obtaining data.

FIGURE 4
Proxy Evidence Gathered from Documentation Reflected on Each Research Poster Showing the Breakdown of the Primary Approach to Obtaining Data

H1. Students who design an experiment and/or collect original data engage in more data practices than students who use external or compiled datasets.

A t-test was used to compare the average total of data practices reported by students whose posters reflected the use of original data with the average total of data practices reported by students whose posters reflected external or compiled datasets (see table 4). The average number of data practices was calculated using the survey results. The data corresponding to the two groups tested, that of the primary data source (that is, original, external, compiled), was calculated independently using the proxy evidence collected from the posters associated

TABLE 4
Comparison of Average Total Data Practices by Students Collecting Original Data vs. Other Research Methods

	Group 1 Original Data Collection	Group 2 Other Research Methods
Mean Data Practices (of 9)	5.143	4.000
Std. Dev.	1.992	1.951
N (# of students)	14	49

with each student (see figure 4). The t-statistic was 1.924528, which was in the 95 percent critical value accepted range (p = .0356). This confirms the hypothesis that students collecting original data engaged in a significantly higher average number of data practices (5.143, n = 14) than students using external or compiled datasets (4.000, n = 49).

H2. Students who use quantitative data analysis techniques engage in more data practices than students who use other techniques to analyze data.

A t-test was used to compare the average total of data practices reported by students whose posters reflected the use of quantitative data analysis techniques with the average total of data practices reported by students whose posters reflected qualitative or other analysis techniques (see table 5). As in the previous hypothesis, the average number of data practices was calculated using the survey results. The data corresponding to the two groups tested, that of the data analysis technique (that is, quantitative, other), was calculated independently using the proxy evidence collected from the posters associated with each student (see figure 3). The t-statistic was 0.62886, which was not in the 95 percent critical value accepted range (p = .531791). This does not support the hypothesis that students conducting quantitative data analysis engaged in a significantly higher average number of data practices (4.364, n = 44) than students using other analysis techniques (4.000, n = 19).

TABLE 5
Comparison of Average Total Data Practices by Quantitative Data Analysis vs. Other Analysis Techniques

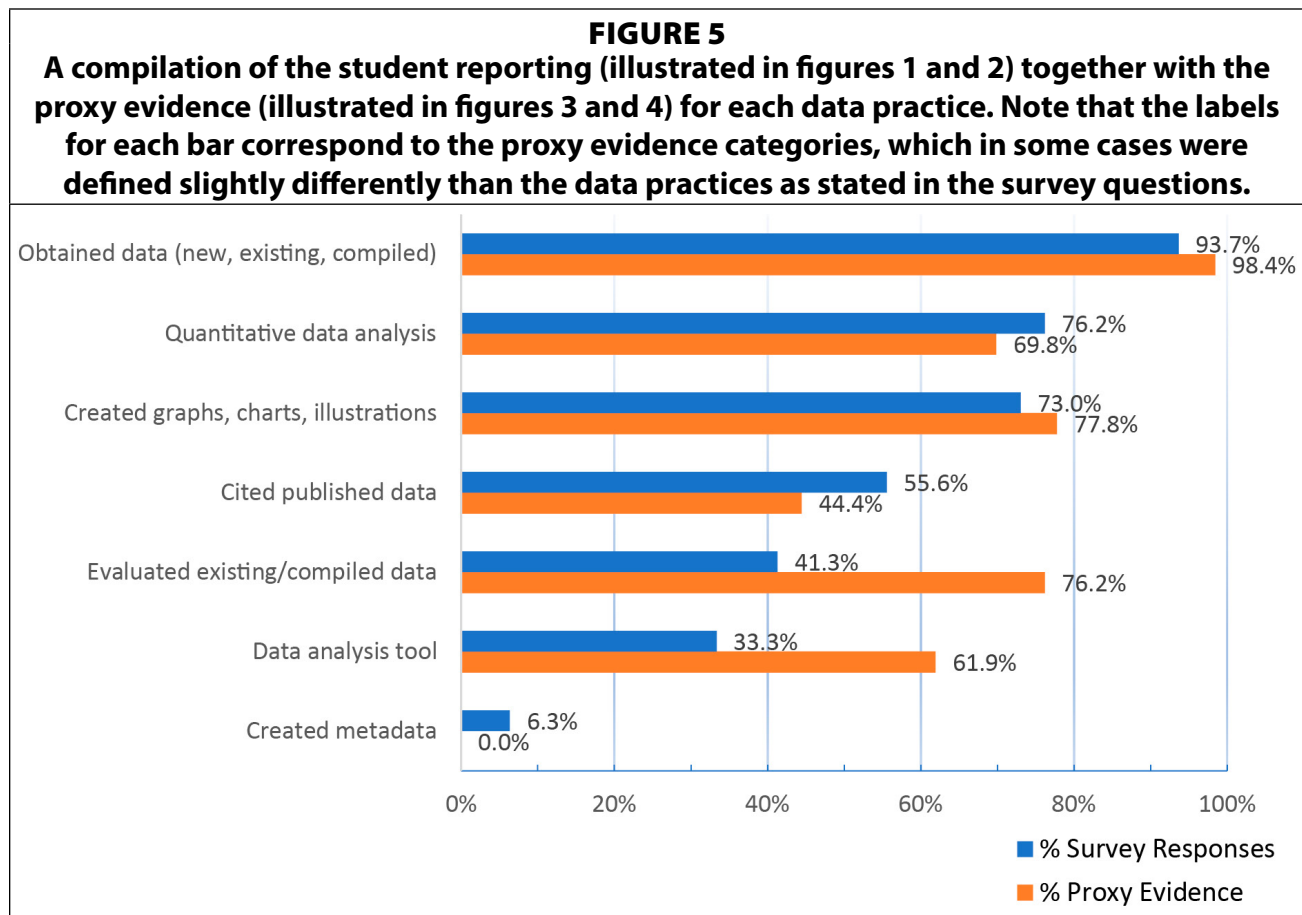|	Group 1 Quantitative Data Analysis	Group 2 Other Techniques
Mean Data Practices (of 9)	4.364	4.000
Std. Dev.	1.810	2.176
N (# of students)	44	19

Discussion
H1 and H2. Student Researcher Engagement with Data Practices

Statistical analysis of the survey results supports the author's hypothesis H1 that students who design an experiment and/or collect original data are likely to report more data practices than students who use external or compiled datasets (see table 4). This finding lends credence to the idea that the creation of an original dataset could be considered a threshold concept that leads to transformative learning with regard to data literacy. Analysis of the hypothesis H2, that students who conduct quantitative data analysis engage with more data practices than students who use other data analysis techniques, was not supported by the data. The lack of a statistically significant result undermines the idea that numerical data analysis is more valid than other data analysis techniques in the practice and development of data literacy skills. This finding aligns with Stanford et al., who found similar gains reported by STEM and non-STEM students after completing early undergraduate research experiences,[34] despite the tendency for research in STEM disciplines to rely more heavily on quantitative data analysis techniques. Further analysis with regard to STEM vs. non-STEM projects, perhaps by sampling a larger population that includes a representative percentage of humanities projects, could

further support both of these assertions.

H3. *Reported Data Practices as Compared with Proxy Evidence of Data Practices*

Careful examination of the research posters and associated application data served to complement and enhance the value of the survey data. The author used this data to test the alignment (see figure 5) between students' understanding of their data practices against the evidence of data practices as reflected on the research posters themselves.

FIGURE 5
A compilation of the student reporting (illustrated in figures 1 and 2) together with the proxy evidence (illustrated in figures 3 and 4) for each data practice. Note that the labels for each bar correspond to the proxy evidence categories, which in some cases were defined slightly differently than the data practices as stated in the survey questions.

In general, the alignment between the survey results and the proxy evidence reflected high levels of agreement with regard to obtaining data and creation of data visualizations (that is, < 5% difference for each data practice). Students reported high levels of participation in each practice, and proxy evidence confirmed this. Hensley's survey showed that only 5 percent of respondents who support formal undergraduate research programs taught searching for statistical information, and even fewer taught data visualization techniques,[35] but the increasing need for this type of library instruction is an opportunity for librarian-faculty curricular development, as illustrated in the case study of collaborative instruction of critical data analysis and visualization by Battista, Boss, and McCartin.[36]

Some of the differences in the percentage agreement between the student reports and the proxy evidence with regard to their approach to obtaining data may be attributed to disci-

plinary differences in understanding of data. For example, the proxy evidence reflected only one student who used no data on their poster, whereas four students (6.3%) did not report obtaining data. An examination of these posters confirmed that only one student performed a true review and synthesis of existing research. This student reported that they had previous experience creating a research poster presentation, they worked alone on this project, and the only data practice they used was citing data sources. Two students who did not report obtaining data used historical methods to analyze primary and secondary sources in projects for anthropology and archaeology courses, respectively. These students did not use numerical data, their posters included a bibliography with several source citations, and both students reported doing the data practice of citing data. One of these students also reported creating data visualizations. The fourth student who did not report obtaining data in the survey actually had done so as confirmed via their research poster narrative, "Data was collected by watching and analyzing videos of ravens, robins, and geese hatching from their eggs." This poster included three data visualizations displaying descriptive statistics of their results. The student reported that they had never previously done a research poster presentation, that they worked alone on this project, and they reported two data practices: 1) analyzing data and 2) creating data visualizations. The poster observation data showed this student's data source to be a compiled dataset, with evidence of quantitative data analysis. It is unclear why this student underreported the extent of their data practices, but it is possible that the student's academic experience led them to believe that video data analysis "doesn't count" as data, despite that they likely recorded their observations initially as text and then converted these into the categorical data displayed in the bar graphs on their poster.

For the data analysis category, agreement between the survey results and proxy evidence was also quite high, with a difference of 6.4 percent. It should be noted that the data analysis category was defined slightly differently in the survey vs. the proxy evidence, which may explain the difference in agreement. The survey asked students whether they "analyzed data," whereas the author specified quantitative analysis when assigning the category to the proxy evidence on each poster. The purpose for this decision was to establish two clear groups for purposes of the H2 analysis.

The misalignment in the remaining categories was considerably higher, which may indicate that students have less understanding of the terminology or the data practice, as in the practices of data citation and cleaning/converting data. As stated earlier, citation of data sources was counted only for those posters that presented external, published data (n = 36), and the proxy evidence showed that less than half of that group (44%) formally cited data sources. However, more than half of survey respondents reported citing their data (n = 63; 55.6%), including the lone student who conducted a review of existing literature and did not present data. This misalignment with the proxy evidence, particularly with regard to citing data sources, may indicate that students don't necessarily discern the differences between "information" and "data" citation as specifically as research faculty or other types of library patrons. Again, Hensley's survey of library instruction for undergraduate research showed that citation management tools were taught by 8.8 percent of respondents.[37] Expanding the scope of this library instruction to include data citation has the potential to benefit students who are new to using data for original research.

The final category reflecting considerable misalignment was that of cleaning/converting data. The proxy evidence reflected that more than 60 percent of students used data analysis

tools, which generally necessitate some cleaning, organizing, and converting data into a specific format for any given analysis tool (such as Excel, JMP, or Tableau). However, only a third of students reported that they had cleaned or converted data.

Analysis of the data show that proxy evidence from the research posters did not fully align with student reports across all data practices. However, student reports of engagement with data practices, together with proxy evidence regarding three data practices (that is, obtaining data, analyzing data, creating visualizations), supports the assertion that data literacy is an integral component of undergraduate research.

Other factors may complicate the alignment between student reports and proxy evidence on the posters themselves and raise new questions for investigation, including the following:

- Did solo researchers engage in more data practices than those who worked collaboratively?
- In projects with faculty co-authors, did the student researchers lack the opportunity to evaluate data, as in the case of students who were provided specific datasets?
- Would the submission of supplemental data files confirm the presence/absence of metadata?

H4. Data Literacy Competencies for Undergraduate Research

The findings from this mixed methods study support the validity and relevance of the undergraduate data literacy competencies established through prior research that collected and analyzed the perspectives of faculty who teach undergraduate courses at two institutions.[38] Through semistructured interviews, this earlier research found that more than 70 percent of faculty agreed that these competencies are relevant for undergraduate education: finding, selecting, accessing, creating datasets; interpreting and evaluating data; communicating data with visualizations; data processing (in this study termed cleaning/converting data); and ethically using/citing data.

The current findings show agreement from the student perspective. More than 70 percent of students reported engaging with the following data practices: finding or creating datasets, analyzing data, and communicating data effectively by creating visualizations. More than half of students reported citing or obtaining permission to use data. While fewer students reported evaluating data, the extensive use of external data sources indicates that the ability to evaluate data sources is an important competency for undergraduate researchers. More than 60 percent of students documented the use of data analysis tools on their posters, which supports the assertion that cleaning and converting data into specific formats is also a data practice used by undergraduate students. This finding in particular supports Carlson's suggestion that undergraduate research could be a point of entry for selected DIL competencies,[39] and Hensley's separate proposal that the "beginner's level" of data-related instruction could be appropriate for library instruction to support undergraduate research programs.[40]

The earlier research on faculty perspectives showed that only 18 percent of faculty saw metadata as an undergraduate activity;[41] this study supports that finding as creating metadata was a data practice reported by only 6.3 percent of respondents (n = 4). A detailed examination of those posters found that two of the four students completed computer modeling projects for biostatistics courses. These students reported engaging in six and all nine data practices, respectively. The third student's poster described a compiled, historical dataset, with visualizations of historical artifacts. The last student's poster reflected a descriptive study in biology that

also used a compiled dataset. For all four students, the author was unable to verify whether the students created metadata based solely on the research posters themselves. The author asserts that the creation of metadata may be applicable for some undergraduate researchers in certain disciplinary circumstances, rather than being widely applicable across disciplines.

Implications for Academic Librarians

For librarians who are responding to institutional needs for integrating data-related topics into their information literacy instruction, this study identifies specific areas of engagement where librarians have expertise. Undergraduate researchers are actively searching for data, evaluating data, and cleaning and/or converting data into formats suitable for creating visualizations. These students are attempting to cite data but don't necessarily have an understanding of how citing data may differ from citing publications.

Data literacy and the practices investigated in this study are embedded throughout the *Framework*. For example, finding and conducting preliminary evaluation of external data sources requires Searching as Strategic Exploration[42] and a recognition that Authority is Constructed and Contextual.[43] Citing external data sources ethically requires an understanding that Information Has Value.[44] Whether a student creates original datasets via the collection or new data or compiles a dataset using multiple external sources, the subsequent analysis, interpretation, and visualization of that data in support of a particular interpretation is an information creation process.[45] Taken together, the data practices investigated in this study are conducted in the broader context of undergraduate research, through which students engage in the knowledge practices described as part of the Research as Inquiry[46] frame.

For a full discussion of the connections between data and information literacy, see Bauder's in-depth treatment of the topic.[47] Bauder argues that data literacy is a natural fit with the *Framework*, in some cases even more so than traditional textual literacy, and highlights ways that Information Creation as a Process and other frames can be incorporated into data literacy instruction.

Conclusion

This study analyzed undergraduate researchers' data practices, triangulating survey results with a detailed examination of the research posters, and found that most students engage in data practices such as obtaining or creating datasets, analyzing data using a variety of techniques and tools, and communicating data through the creation of data visualizations. While disciplines that use experimental design and original data collection may require a broader range of data literacy competencies, a substantial number of data practices are used in undergraduate research across disciplines.

Effective data literacy instruction in support of undergraduate research, whether it is course-based or a formal program, has the potential to become a primary way to improve the data literacy of graduating students, thereby preparing them to join today's data-driven workforce. The language of the *Framework* provides a flexible toolkit that positions academic librarians to play a pivotal role, whether by developing curriculum in partnership with faculty or by developing co-curricular programming in partnership with Offices of Research or other symposia organizers.

To that end, this study lays the groundwork for additional research. The author intends to analyze the survey data with regard to the data-related challenges students experienced

during the course of their research and the ways in which they used library resources. Further analysis may clarify whether library interventions such as self-guided online tutorials, videos, and other instruction would be beneficial in areas such as data cleaning, processing, citation, and data ethics and integrity.

A future study will further explore and identify suitable proxy evidence from undergraduate research work products including posters and other supplemental files such as audio or video presentations, supplemental data files, and the like. Further investigation in this area could inform the development of assessment tools that would be feasible for institutionwide assessments of data literacy.

Acknowledgments

Notes

1. Roger S. Rowlett, Linda Blockus, and Susan Larson, "Characteristics of Excellence in Undergraduate Research (COEUR)," in *Characteristics of Excellence in Undergraduate Research*, ed. Nancy Hensel (Washington, DC: Council of Undergraduate Research, 2012), 2–19.

2. Javier Calzada Prado and Miguel Ángel Marzal, "Incorporating Data Literacy into Information Literacy Programs: Core Competencies and Contents," *Libri* 63, no. 2 (2013): 123–34, https://doi.org/10.1515/libri-2013-0010.

3. Association of College and Research Libraries (ACRL), *Framework for Information Literacy for Higher Education* (Chicago, IL: Association of College and Research Libraries, 2016), www.ala.org/acrl/sites/ala.org.acrl/files/content/issues/infolit/Framework_ILHE.pdf.

4. David Lopatto, "Adapting to Change: Studying Undergraduate Research in the Current Education Environment," *Scholarship and Practice of Undergraduate Research* 1, no. 1 (2017): 5–10, https://doi.org/10.18833/spur/1/1/7.

5. Lee Bracher, Jane Cantrell, and Kay Wilkie, "The Process of Poster Presentation: A Valuable Learning Experience," *Medical Teacher* 20, no. 6 (1998): 552–57, https://doi.org/10.1080/01421599880274.

6. Theresa Burress, Emily Mann, and Tina Neville, "Exploring Data Literacy via a Librarian-Faculty Learning Community: A Case Study," *Journal of Academic Librarianship* 46, no. 1 (2020): 102076, https://doi.org/10.1016/j.acalib.2019.102076.

7. Theresa Burress et al., "Data Literacy in Undergraduate Education: Faculty Perspectives and Pedagogical Approaches," in *Data Literacy in Academic Libraries: Teaching Critical Thinking with Numbers*, ed. Julia Bauder (Chicago, IL: American Library Association, 2021), 1–22.

8. Rowlett, Blockus, and Larson, "Characteristics of Excellence in Undergraduate Research (COEUR)."

9. George D. Kuh, *High-Impact Educational Practices: What They Are, Who Has Access to Them, and Why They Matter, Leap* (Washington, DC: Association of American Colleges and Universities, 2008), https://doi.org/10.1017/CBO9781107415324.004.

10. Lopatto, "Adapting to Change."

11. Bracher, Cantrell, and Wilkie, "The Process of Poster Presentation."

12. Sharyn J. Potter et al., "Intellectual Growth for Undergraduate Students: Evaluation Results from an Undergraduate Research Conference," *Journal of College Teaching & Learning* 7, no. 2 (2010): 25–34, https://doi.

org/10.19030/tlc.v7i2.86.

13. Jennifer S. Stanford et al., "Early Undergraduate Research Experiences Lead to Similar Learning Gains for STEM and Non-STEM Undergraduates," *Studies in Higher Education* 42, no. 1 (2017): 115–29, https://doi.org/10.1080/03075079.2015.1035248.

14. Adrianna Kezar and Elizabeth Holcombe, "Support for High-Impact Practices: A New Tool for Administrators," *Liberal Education* 103, no. 1 (2017): 34–39.

15. Lynn Silipigni Connaway et al., *Academic Library Impact: Improving Practice and Essential Areas to Research* (Chicago, IL: Association of College and Research Libraries, 2017), http://www.ala.org/acrl/sites/ala.org.acrl/files/content/publications/whitepapers/academiclib.pdf.

16. Rita Vine, "Realigning Liaison with University Priorities," *College & Research Libraries News* 79, no. 8 (2018): 420–23, 458.

17. Rowlett, Blockus, and Larson, "Characteristics of Excellence in Undergraduate Research (COEUR)."

18. ACRL, *Framework for Information Literacy for Higher Education*.

19. Merinda Kaye Hensley, Sarah L. Shreeves, and Stephanie Davis-Kahl, "A Survey of Library Support for Formal Undergraduate Research Programs," *College & Research Libraries* 75, no. 4 (2014): 422–41, https://doi.org/10.5860/crl.75.4.422.

20. Merinda Kaye Hensley, "A Survey of Instructional Support for Undergraduate Research Programs," *portal: Libraries & the Academy* 15, no. 4 (2015): 719–46, https://doi.org/10.5860/crl.76.7.975.

21. *Data Information Literacy: Librarians, Data, and the Education of a New Generation of Researchers*, eds. Jake Carlson and Lisa R. Johnston (West Lafayette, IN: Purdue University Press, 2015), http://www.datainfolit.org/dilguide/.

22. Jake Carlson et al., "Determine Data Information Literacy Needs: A Study of Students and Research Faculty," in *Data Information Literacy: Librarians, Data, and the Education of a New Generation of Researchers*, eds. Jake Carlson and Lisa R. Johnston (West Lafayette, IN: Purdue University Press, 2015), 11–33.

23. Jake Carlson, "Future Directions for Data Information Literacy," in *Data Information Literacy: Librarians, Data, and the Education of a New Generation of Researchers*, eds. Jake Carlson and Lisa R. Johnston (West Lafayette, IN: Purdue University Press, 2015), 247–60; Yasmeen Shorish, "Data Information Literacy and Undergraduates: A Critical Competency," *College and Undergraduate Libraries* 22, no. 1 (2015): 97–106, https://doi.org/10.1080/10691316.2015.1001246.

24. Andrew Battista, Katherine Boss, and Marybeth McCartin, "Data Literacy in Media Studies: Strategies for Collaborative Teaching of Critical Data Analysis and Visualization," *Journal of Interactive Technology and Pedagogy* 18 (2020), https://jitp.commons.gc.cuny.edu/data-literacy-in-media-studies-strategies-for-collaborative-teaching-of-critical-data-analysis-and-visualization/.

25. *Data Literacy in Academic Libraries: Teaching Critical Thinking with Numbers*, ed. Julia Bauder (Chicago, IL: American Library Association, 2021).

26. Burress, Mann, and Neville, "Exploring Data Literacy via a Librarian-Faculty Learning Community."

27. Prado and Marzal, "Incorporating Data Literacy into Information Literacy Programs."

28. Burress, Mann, and Neville, "Exploring Data Literacy via a Librarian-Faculty Learning Community."

29. Burress et al., "Data Literacy in Undergraduate Education."

30. George D. Kuh and Ken O'Donnell, "High-Impact Practices: Eight Key Elements and Examples," in *Ensuring Quality & Taking High-Impact Practices to Scale* (Washington, DC: Association of American Colleges & Universities, 2013).

31. Patricia Leavy, *Research Design: Quantitative, Qualitative, Mixed Methods, Arts-Based, and Community-Based Participatory Research Approaches* (New York, NY: Guilford Publications, 2017), http://ebookcentral.proquest.com/lib/usf/detail.action?docID=4832778.

32. Burress et al., "Data Literacy in Undergraduate Education"; Burress, Mann, and Neville, "Exploring Data Literacy via a Librarian-Faculty Learning Community."

33. Paige Morgan, "This Talk Doesn't Have a Name," *Paige Morgan Blog* (2017), http://blog.paigemorgan.net/articles/17/this-talk.html.

34. Stanford et al., "Early Undergraduate Research Experiences Lead to Similar Learning Gains for STEM and Non-STEM Undergraduates."

35. Hensley, "A Survey of Instructional Support for Undergraduate Research Programs."

36. Battista, Boss, and McCartin, "Data Literacy in Media Studies: Strategies for Collaborative Teaching of Critical Data Analysis and Visualization."

37. Hensley, "A Survey of Instructional Support for Undergraduate Research Programs."

38. Burress et al., "Data Literacy in Undergraduate Education."

39. Carlson, "Future Directions for Data Information Literacy."

40. Hensley, "A Survey of Instructional Support for Undergraduate Research Programs."

41. Burress et al., "Data Literacy in Undergraduate Education."

42. ACRL, *Framework for Information Literacy for Higher Education*, 22.

43. ACRL, *Framework for Information Literacy for Higher Education*, 12.

44. ACRL, *Framework for Information Literacy for Higher Education*, 16.

45. ACRL, *Framework for Information Literacy for Higher Education*, 14; Battista, Boss, and McCartin, "Data Literacy in Media Studies."

46. ACRL, *Framework for Information Literacy for Higher Education*, 18.

47. Bauder, *Data Literacy in Academic Libraries.*