Data literacy in undergraduate research: A case study from student poster competitions

Margaret Marchant[1] and C. Jeffrey Belliston[2]

Abstract

At universities, research involving data is often regarded as the domain of graduate students and faculty. However, undergraduate students also work with data within the research process, and it can be a core experience to prepare them for future education and careers. Research products from undergraduate students can demonstrate the extent of their data literacy skills and understanding, which are becoming central to success in graduate studies and the world of work. Since a leading way for undergraduate students to share research is through posters, this paper examines undergraduate posters at Brigham Young University (BYU) in the context of data literacy skills. The paper defines data literacy and the importance of undergraduate students becoming data literate. This case study shares the BYU context for the undergraduate poster competitions and the resulting strengths and gaps in data literacy education followed by suggestions for supporting and encouraging undergraduate research and data literacy development beyond the traditional area of data analysis.

Keywords

data literacy, undergraduates, posters, case study

Introduction

In today's collaborative scientific enterprise, data 'have become more valuable as . . . [stand-alone] scholarly product[s] with potential for reuse' (Shorish, 2015, p. 98). As research across disciplines becomes more data-focused, it is essential for researchers, including student researchers, to be data literate. Although there is not yet a standard definition for data literacy, there is much consensus among researchers about the main elements of data literacy, including areas of describing, cleaning, analyzing, evaluating, storing, and sharing data, which the current study uses to define data competency categories (Carlson et al., 2011; Calzada Prado and Marzal, 2013).

The need for undergraduates to obtain data literacy competency is true regardless of whether they pursue graduate study or enter the labor force. In her article describing data information literacy as 'a critical competency' for undergraduates, Shorish (2015) stated, 'Therefore, as one seeks to create a more informed and productive citizenry, one should seek to expose all college graduates to the skills required to effectively evaluate and use data' (p. 102).

Undergraduates studying social sciences and life sciences disciplines at Brigham Young University (BYU) in Provo, Utah, USA, have been creating posters sharing research results for over a decade. This trove of products was analyzed using both manual and automated content analysis methods to identify skills and gaps in data literacy competencies among undergraduate researchers. The longitudinal and multi-disciplinary nature of the poster archives allowed for the analysis of changes over time and across fields.

Specifically, we sought to address 1) how and what kind of data undergraduate students use in their research; 2) what data literacy skills and gaps undergraduates exhibit in their research; 3) heterogeneous effects based on the discipline, analytical methodology, type of data, or experience of the student researchers. The findings guide university faculty, librarians, and others in mentoring undergraduate students in research. Students already possess many skills they can harness in future educational and career endeavors. Targeted support in weak areas will prepare students for a world ever more saturated with data.

Literature Review

Data literacy is the ability to find, interpret, analyze, and communicate with and about data. A recent review of the literature on data literacy education demonstrated that there is room for more research (Ghodoosi, 2023). Our study helps to fill this gap by exposing strengths and weaknesses in students' data skills, particularly in their ability to communicate about data.

Undergraduate Research

Research within institutions of higher education has traditionally been viewed as the domain of faculty and graduate students. However, universities are increasingly recognizing the benefit of research experiences for undergraduate students. Experiential learning related to research is a high-impact practice because it is particularly effective in helping students develop skills and provides added benefits for students from traditionally underrepresented groups (American Association of Colleges and Universities, 2024).

Researchers have assessed the positive impact of undergraduate research through multiple modes, including course-integrated, applied research projects (Stark et al., 2018; Pratoomchat and Mahjabeen, 2023), data fellowships (Carter, 2021), mentored research (Gilmore et al., 2015; Ruth et al., 2023), class poster sessions (Stegemann and Sutton-Brady, 2009; Kinikin and Hench, 2012; Altintas et al., 2014; Logan, Quiñones, and Sunderland, 2015; Duckworth and Halliwell, 2022), or research conference poster sessions (Mabrouk, 2009; Burress, 2022). Students benefit from these experiences by being more engaged and self-motivated in learning (Mabrouk, 2009; Stegemann and Sutton-Brady, 2009), reducing learning anxiety (Stegemann and Sutton-Brady, 2009), improving comprehension of concepts (Kinikin and Hench, 2012; Altintas et al., 2014), and gaining transferable skills for graduate school and the workplace (Gilmore et al., 2015; Carter, 2021; Duckworth and Halliwell, 2022; Pratoomchat and Mahjabeen, 2023; Ruth et al., 2023). The most significant of these skills include learning to collect, analyze, visualize, communicate, and manage data.

The benefits of undergraduate research relating to data literacy skills are most relevant to our study. A study of implementing a real-world research project into undergraduate economics principles courses found that 90% of students agreed that participating in the research improved their skill in collecting, processing, and interpreting data (Pratoomchat and Mahjabeen, 2023). Similarly, Ruth et al. (2023) assessed a new undergraduate research program in the social sciences. Findings from surveying undergraduate students and their research mentors pointed to data management as one of the major skills gained, along with collecting and analyzing data.

Since one of the most common and most accessible forms of sharing undergraduate research experiences is a poster session, this study focuses on that mode. Posters have the added benefit of being a product that can be shared and evaluated over time to assess students' data literacy.

Research Posters and Data Literacy

To date, there have been a few studies on undergraduate research posters that connect to students' data literacy skills. Duckworth and Halliwell's (2022) content analysis of 100 posters from a multidisciplinary virtual poster session found that students are most comfortable with presenting data visually (73%) or through reporting results of data analysis (70%) and are less comfortable evaluating sources, including data sources (42%). This corroborates Logan, Quiñones, and Sunderland whose 2015 report of a longitudinal study of a poster presentation project in a lower-level chemistry course found that data visualization does not come naturally to students, that training on poster and figure design improved the posters, and that having the opportunity to present their findings in this format contributed to students' feelings of efficacy in their learning.

Burress's (2022) evaluation of 58 undergraduate posters in multiple fields used both student self-reports of the use of data practices and proxy evidence from content analysis of the posters. The study focused on visualization, analysis, evaluation, citation, cleaning, and metadata creation. While most students (98%) reported using at least one of these data practices content analysis of the posters did not support their reports. Burress suggests this could be due to a lack of understanding of academic terminology and called for consideration of additional proxy evidence from research posters, which our study provides.

Our research adds to previous studies by evaluating a more comprehensive sample of student work. We compare student research from different disciplines (social sciences and life sciences) over multiple years. A larger sample size allows us to produce more precise estimates of areas of strength and weakness in student data literacy. Since approaches for studying text and image data from research posters are inherently more subjective, it is useful to have multiple studies in different contexts and methods to understand what findings are reproducible and generalizable. We also address findings on students' lack of knowledge of academic terminology by using an automated text search in addition to manual content analysis.

Brigham Young University

Brigham Young University (BYU) is a private, faith-based university sponsored by The Church of Jesus Christ of Latter-day Saints. BYU has the dual mission of strengthening students' faith in the Lord Jesus Christ and providing undergraduates the opportunity to gain a rigorous education in one or more of 198 majors and 113 minors. Though having a Research 1 Carnegie classification (very high research spending and doctorate production), BYU is primarily an 'undergraduate teaching institution.'[3] Slightly more than 32,000 undergraduates comprise the vast majority of the student body of about 35,000 daytime students.

Undergraduate Research at BYU

Building on the work of past leadership, current BYU President, C. Shane Reese chose *strengthening the student experience* as his top strategic goal. In his inaugural remarks he stated, "becoming BYU will require enriching the student experience and strengthening our already student-centric approach" (Tanner, 2024, p. 306).

Accordingly, there is a strong emphasis placed on the involvement of undergraduate students in research conducted by BYU faculty. Evidence of this is shown in the attention given to student mentoring in BYU's *Rank and Status Policy*. All BYU faculty are expected to mentor students and encouraged specifically that, "involving and mentoring students in high-quality scholarship can deepen their learning and expand future opportunities" (BYU, 2022b, section 3.3). The procedural documents that guide implementation of the Rank and Status policy also include student mentorship (where possible) as a criterion for evaluating scholarship (BYU, 2022c; BYU, 2022d).

Moving toward concrete and measurable learning outcomes, BYU administration encourages its constituent colleges to develop specific outcomes from mentored undergraduate research (e.g., becoming data literate) rather than developing university-wide outcomes. The University provides significant resources to colleges to help achieve the research learning outcomes they set forth (Howell, 2024, March).

Over the 2013–2022 calendar years, BYU's receipt of external research funds averaged $35.7 million USD. Of the $39.4 million of external research funds received in 2022, nearly 25% ($9.7 million) were directed to supporting students involved in research. That same year, the University directly supported experiential learning to the tune of nearly $4.1 million, and colleges kicked in an additional $11.2 million for total experiential learning support of $25 million. In the most recent year, experiential learning financial support from all sources increased from just over $25 million USD to $37.3 million spread over 18,207 student experiences (BYU, 2022a; Howell, 2024, August).

In addition to structural support of mentored undergraduate research, leaders at BYU frequently provide meaningful rhetorical encouragement. Core elements of the university mission include commitment to excellence in research and the development of the full potential of students, which are interwoven in mentored undergraduate research (Worthen, 2017). The deep institutional support of undergraduate research, both financially and ideologically, sets the stage for our study of the research works produced by students under faculty mentorship.

BYU Poster Sessions

Individual colleges at BYU host opportunities for undergraduates to share their research. Poster competitions where the posters are archived enable us to study data literacy skills after the fact. The FHSS Mentored Research Conference[4] is run every year by BYU's College of Family, Home, and Social Sciences (FHSS) to support undergraduate mentored research experiences and give students opportunities to practice sharing their research. The goal[5] is to prepare students for future careers and/or grad school. The Library/Life Sciences Undergraduate Poster Competition[6] is co-hosted by the BYU Library and the College of Life Sciences. The goal[7] of this poster session is to help students learn and practice communicating research to general audiences (Frost, Goates, and Nelson, 2023).

Methodology

Extending from past research (Logan, Quiñones, and Sunderland, 2015; Burress, 2022; Duckworth and Halliwell, 2022), this study uses student research posters as the data source to study students' data literacy competencies, gaps in skills, and differences across disciplines and over time. The benefit of using student posters is to examine what students do in practice and how they communicate their work with and understanding of data. This study adds to previous research by utilizing a large sample of student research posters archived over more than 10 years and comprising both social sciences and life sciences. This provides a greater sample to validate trends and differences across groups.

Sample

The data comes from two collections of student posters in Brigham Young University's ScholarsArchive repository. The FHSS Mentored Research Conference archives 240 student posters from social science disciplines published from 2010 to 2023. The Library/Life Sciences Undergraduate Poster Competition has grown each year since it started in 2017. As of the end of 2023, there were 213 archived posters. Some of the posters (N = 23) in the collections provided overviews of a topic but did not report on a research study. Additionally, a few posters were created by graduate students (N = 9). Since these posters were out of the research question scope, they were excluded from the final analysis, making the total sample size 421 posters.

The class standing of students submitting posters to the competitions ranged from freshmen to seniors. Table 1 reports the proportion of students in each year of school who submitted posters. The online submission form for the Life Sciences poster competition included information about class standing, which is not a required metadata field in ScholarsArchive. Thus, the class standing data is more complete for the Life Science posters. Where student class standing information was available, most students creating research posters were juniors and seniors. The majors of students submitting posters ranged from Biology to Neuroscience within the Life Sciences and from Anthropology to Sociology in the Social Sciences. The most common life science major was Biology (N = 62). The most common social science majors were Psychology (N = 52) and Family Life (N = 50).

Table 1. Sample Averages for Demographic and Data Literacy Characteristics of Undergraduate Research Posters, by Poster Discipline

	All Posters	Social Sciences	Life Sciences
Freshmen	0.02 (0.13)	0*** (0)	0.03 (0.18)
Sophomore	0.05 (0.21)	0.01*** (0.10)	0.09 (0.28)
Junior	0.17 (0.38)	0.03*** (0.17)	0.31 (0.46)
Senior	0.32 (0.47)	0.10*** (0.30)	0.53 (0.50)
Unspecified class standing	0.45 (0.50)	0.86*** (0.35)	0.04 (0.19)

Sample	421	210	211

Notes: Data from undergraduate research posters archived in BYU ScholarsArchive for the FHSS Mentored Research Conference and Life Sciences Competition. T-tests were run to compare the outcomes of the social sciences and life sciences posters. * p < .10, ** p < .05, *** p < .01.

Data Collection

Six main data literacy competency areas were identified: describing data, cleaning data, analyzing data, evaluating data, sharing data, and storing data. Proxy data to measure the presence of each competency was collected using two different content analysis methods. This provides multiple perspectives to view each competency and increases the reliability of the results.

First, a manual content analysis of each poster was performed. Marchant defined criteria for measuring each data literacy competency category (see Appendix A for the full definitions). Data was collected by a research assistant reviewing each poster in relation to the defined criteria. Before data for all posters was collected, the author and research assistant each coded 20 posters and compared responses. The coding definitions were adjusted as necessary to align the data. After all poster data was collected, a random selection of 30% of the posters was verified by Marchant to ensure consistency and accuracy.

Second, data on term frequencies was collected using the automated Adobe Acrobat Index Search feature. Marchant selected common terms relating to each phase of the research data process. This was accomplished by 1) close reading of 20 posters (ten each from the social sciences and life sciences poster competitions), with the intent to identify terms used when discussing data and data practices, and 2) research within the sphere of data literacy to identify other terms used to describe working with data (Carlson et al., 2011; Calzada Prado and Marzal, 2013). Repeated terms were added to the search term list. Appendix B lists the terms along with the related data practice category. The proportion of posters using each term[8] was identified. This provides a more objective measure of which data practices student researchers participate in.

Results

Table 2 reports the descriptive statistics for variables collected manually from the posters. In both poster competitions, there was an emphasis on quantitative data and research methods and primary data collection. Of the sections on the poster that relate to data, the most used was a results section (77%), followed by data descriptions (47%). Data was visualized in a variety of tabular and graphical ways, with data representation in figures being more widespread.

Table 2. Sample Averages for Data Literacy Characteristics of Undergraduate Research Posters, by Poster Discipline

	All Posters	Social Sciences	Life Sciences
Methods	0.82 (0.39)	0.80 (0.40)	0.84 (0.37)
Used quantitative method	0.73 (0.44)	0.66*** (0.48)	0.80 (0.40)

Used qualitative method	0.15 (0.35)	0.20*** (0.40)	0.09 (0.29)
Used mixed methods	0.12 (0.32)	0.14* (0.35)	0.09 (0.29)
Used primary data	0.71 (0.45)	0.55*** (0.50)	0.88 (0.33)
Used secondary data	0.30 (0.46)	0.46*** (0.50)	0.15 (0.36)
Describe data	0.47 (0.50)	0.68*** (0.47)	0.27 (0.44)
Analyze data	0.77 (0.42)	0.75 (0.43)	0.78 (0.41)
Clean data	0.20 (0.40)	0.21 (0.41)	0.19 (0.39)
Evaluate data	0.11 (0.32)	0.17*** (0.38)	0.06 (0.23)
Include form data citation	0.07 (0.25)	0.08 (0.27)	0.06 (0.23)
Share Data	0.01 (0.10)	0.01 (0.10)	0.01 (0.10)
Store Data	0 (0)	0 (0)	0 (0)
Number of data tables	0.61 (1.03)	0.91*** (1.19)	0.31 (0.73)
Number of data visualizations	3.03 (2.30)	2.03*** (1.97)	4.03 (2.16)
Sample	421	210	211

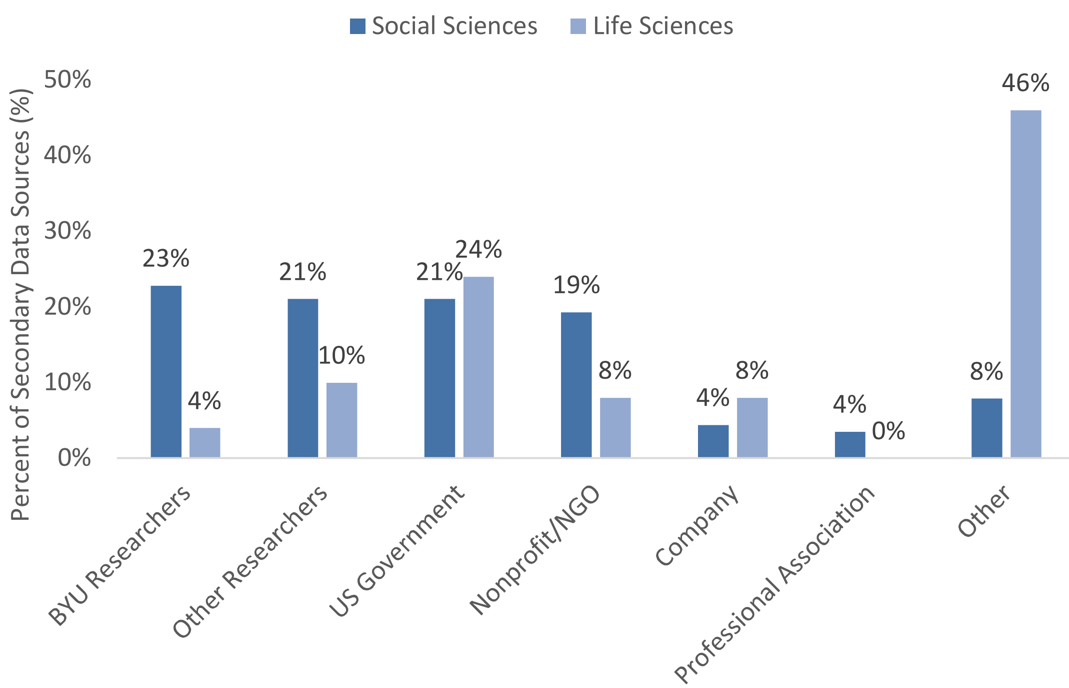
Notes: Data from undergraduate research posters archived in BYU ScholarsArchive for the FHSS Mentored Research Conference and Life Sciences Competition. T-tests were run to compare the outcomes of the social sciences and life sciences posters. * $p < .10$, ** $p < .05$, *** $p < .01$.

T-tests were also performed comparing posters from the social sciences and life sciences. Several significant differences between the two poster competitions highlight differences in logistics and more meaningful, unique characteristics and research patterns. The Life Science competition incorporates a robust Qualtrics survey for collecting metadata on the poster and student creator, which likely helped to provide more consistent coverage of student class standing.

The differences in the type of methodology and data used in research between the social sciences and life sciences posters point to differences in research approach between the two domains. While still quantitatively focused, social science research posters showed a greater variety of research and data used, including qualitative and mixed methods and secondary data. Students doing research in the social sciences were also more likely to include a data description and discuss the quality of the research data. The more open display of these competencies may be connected to the different types of data used since students may feel more need to describe and justify their choice to use an outside data source.

As we explored the specific sources of secondary data used by social science and life science undergraduate researchers, we identified common sources and types of data used. There were 81 unique data sources used in the social sciences research posters and 35 unique data sources used in the life sciences posters, reflecting the greater percentage of posters using secondary data in the social sciences. As shown in Figure 1, students researching in the social sciences more often used data collected by other researchers (either at BYU or other institutions). Social science posters also used nonprofit or NGO data more frequently.

Figure 1. Type of Secondary Data Used in Undergraduate Research Posters, by Poster Discipline



Notes: Data from undergraduate research posters archived in BYU ScholarsArchive for the FHSS Mentored Research Conference and Life Sciences Competition.

Table 3 reports the proportion of articles including each term in the context of using or communicating data. Overall, students most used terms related to describing (73%), analyzing (69%), and evaluating (71%) data. A minority of students (19%) used terms related to data management, either sharing or storing data. These results on describing, analyzing, sharing, and storing data align with what we found through the manual coding of posters. The large difference in the evaluating data metrics (11% for manual coding and 71% for automated indexing) indicates the complexity of measuring, teaching, and demonstrating this competency. It also suggests that students may be familiar with terms related to evaluating data (e.g., compare or limitations) but not yet be able to clearly communicate the data quality or how they evaluated it.

Table 3. Proportion of Articles with Data Literacy Competency Terms, by Poster Discipline

Term	All	Social Sciences	Life Sciences
Describing	0.73 (0.44)	0.73 (0.44)	0.73 (0.44)
Average	0.29 (0.45)	0.28 (0.45)	0.29 (0.45)
Binary	0.03 (0.18)	0.04 (0.20)	0.02 (0.15)
Categoric*	0.01 (0.08)	0.01* (0.12)	0 (0)
Continuous	0.03 (0.17)	0.03 (0.17)	0.03 (0.17)
Discrete	0.01 (0.10)	0.01 (0.10)	0.01 (0.10)
Frequency	0.13 (0.34)	0.12 (0.32)	0.14 (0.35)
Longitudinal	0.10 (0.30)	0.18*** (0.38)	0.03 (0.17)
Mean	0.34 (0.48)	0.40** (0.49)	0.29 (0.45)
Median	0.03 (0.18)	0.02 (0.15)	0.04 (0.20)
Nominal	0 (0)	0 (0)	0 (0)
Ordinal	0.01 (0.08)	0.01* (0.12)	0 (0)
Population	0.26 (0.44)	0.26 (0.44)	0.26 (0.44)
Proportion	0.07 (0.25)	0.06 (0.23)	0.08 (0.27)
Random	0.14 (0.35)	0.14 (0.35)	0.13 (0.34)
Standard deviation	0.04 (0.19)	0.05 (0.21)	0.02 (0.15)
Subjects	0.11 (0.31)	0.13* (0.34)	0.08 (0.27)
Cleaning	0.39 (0.49)	0.31*** (0.47)	0.46 (0.50)
Calculat*	0.13 (0.34)	0.09*** (0.28)	0.18 (0.39)
Clean	0.02 (0.15)	0.02 (0.14)	0.03 (0.17)

Construct	0.09 (0.29)	0.11* (0.32)	0.07 (0.25)
Conver*	0.08 (0.28)	0.08 (0.27)	0.09 (0.28)
Extract	0.06 (0.24)	0.01*** (0.12)	0.10 (0.31)
Merge	0.07 (0.26)	0.06 (0.24)	0.08 (0.27)
Normalize	0.05 (0.23)	0.02*** (0.14)	0.09 (0.29)
Analyzing	0.69 (0.46)	0.72 (0.45)	0.66 (0.47)
Analy*	0.57 (0.50)	0.60 (0.49)	0.54 (0.50)
Correlat*	0.27 (0.44)	0.31* (0.46)	0.23 (0.42)
P-value	0.05 (0.22)	0.05 (0.22)	0.05 (0.22)
Regress*	0.19 (0.39)	0.25*** (0.44)	0.12 (0.32)
T-test	0.04 (0.20)	0.05 (0.21)	0.03 (0.18)
Variation	0.09 (0.29)	0.09 (0.28)	0.10 (0.30)
Evaluating	0.71 (0.46)	0.69 (0.47)	0.73 (0.45)
Accura*	0.10 (0.30)	0.07* (0.26)	0.12 (0.33)
Appropriate	0.05 (0.21)	0.08*** (0.27)	0.01 (0.10)
Authority	0.01 (0.08)	0.01* (0.12)	0 (0)
Compar*	0.47 (0.50)	0.37*** (0.48)	0.57 (0.50)
Credib*	0.00 (0.07)	0.00 (0.07)	0.00 (0.07)
Limi*	0.31 (0.46)	0.34 (0.48)	0.27 (0.45)
Quality	0.20 (0.40)	0.25*** (0.43)	0.15 (0.35)
Reliab*	0.05 (0.23)	0.06 (0.24)	0.05 (0.21)
Sharing	0.19	0.18	0.19

	(0.39)	(0.38)	(0.40)
Availabl*	0.13 (0.34)	0.13 (0.34)	0.13 (0.33)
Replicat*	0.06 (0.23)	0.04 (0.20)	0.07 (0.26)
Repository	0.00 (0.05)	0 (0)	0.00 (0.07)
Storing	0.19 (0.40)	0.19 (0.39)	0.20 (0.40)
Archiv*	0.03 (0.16)	0.03 (0.18)	0.02 (0.14)
Confidential	0 (0)	0 (0)	0 (0)
Manag*	0.11 (0.31)	0.10 (0.29)	0.12 (0.32)
Preserv*	0.03 (0.17)	0.01** (0.12)	0.05 (0.21)
Secur*	0.05 (0.23)	0.07 (0.26)	0.04 (0.19)
Sample	421	210	211

Notes: Data from undergraduate research posters archived in BYU ScholarsArchive for the FHSS Mentored Research Conference and Life Sciences Competition. T-tests were run to compare the outcomes between social sciences and life sciences posters. * $p < .10$, ** $p < .05$, *** $p < .01$.

When comparing the proportions between social sciences and life sciences posters, the only statistically significant difference in a major data literacy competency category was for data cleaning. Students doing life science research posters discussed how they cleaned or processed their data before data analysis more (46% compared to 31%). Statistically significant differences between proportions of individual terms suggest that this comes from more discussion of calculating and normalizing variables and extracting data, which all may be more common when collecting primary data.

Burress's (2022) findings pointed to differences in data literacy competencies between students who conducted quantitative research or collected their own data. We tested these relationships through two-sample t-tests, comparing posters that used quantitative methods with those that used qualitative or mixed methods (see Table 4) and comparing posters that used primary data with those that used secondary data (see Table 5). We found that posters that reported on quantitative research were more likely to include method, data description, and results sections on the poster compared to qualitative or mixed method studies. Unsurprisingly, posters reporting on quantitative research also included more tables and figures. This suggests that students conducting quantitative research have more practice with skills of describing, analyzing, and visualizing data. On the other hand, posters that reported on research with primary data were less likely to include data descriptions or mention cleaning data. Secondary data users more often cited their data than primary data users.

Table 4. Proportion of Articles with Data Literacy Competency Practice, by Method

	Quantitative	Qualitative or Mixed Methods
Manually Coded Measures		
Methods	0.84* (0.37)	0.76 (0.43)
Used primary data	0.72 (0.45)	0.69 (0.46)
Used secondary data	0.30 (0.46)	0.31 (0.46)
Describe data	0.50** (0.50)	0.38 (0.49)
Analyze data	0.82*** (0.38)	0.62 (0.49)
Clean data	0.21 (0.41)	0.17 (0.37)
Evaluate data	0.13* (0.34)	0.07 (0.26)
Include form data citation	0.08 (0.27)	0.04 (0.21)
Share Data	0.01 (0.11)	0 (0)
Store Data	0 (0)	0 (0)
Number of data tables	0.68** (1.02)	0.42 (1.03)
Number of data visualizations	3.25*** (2.37)	2.45 (1.96)
Automated Indexing Measures		
Describing Data	0.78*** (0.42)	0.61 (0.49)
Cleaning Data	0.40 (0.49)	0.35 (0.48)
Analyzing Data	0.72** (0.45)	0.61 (0.49)
Evaluating Data	0.71 (0.45)	0.68 (0.47)
Sharing Data	0.20 (0.40)	0.15 (0.36)
Storing Data	0.18 (0.38)	0.24 (0.43)
Sample	307	114

Notes: Data from undergraduate research posters archived in BYU ScholarsArchive for the FHSS Mentored Research Conference and Life Sciences Competition. T-tests were run to compare the outcomes between posters using a quantitative method to those with a qualitative or mixed method. * p < .10, ** p < .05, *** p < .01.

Table 5. Proportion of Articles with Data Literacy Competency Practice by Type of Data

	Primary Data	Secondary Data
Manually Coded Measures		
Methods	0.86*** (0.35)	0.72 (0.45)
Used quantitative method	0.74 (0.44)	0.71 (0.46)
Used qualitative method	0.14 (0.35)	0.16 (0.37)
Used mixed methods	0.12 (0.33)	0.11 (0.31)
Describe data	0.44** (0.50)	0.55 (0.50)
Analyze data	0.77 (0.42)	0.77 (0.42)
Clean data	0.16*** (0.37)	0.29 (0.46)
Evaluate data	0.11 (0.31)	0.13 (0.34)
Include form data citation	0.02*** (0.15)	0.17 (0.37)
Share Data	0.01 (0.08)	0.02 (0.13)
Store Data	0 (0)	0 (0)
Number of data tables	0.52*** (0.99)	0.85 (1.10)
Number of data visualizations	3.40*** (2.20)	1.93 (1.90)
Automated Indexing Measures		
Describing Data	0.74 (0.44)	0.70 (0.46)
Cleaning Data	0.40 (0.49)	0.36 (0.48)
Analyzing Data	0.67 (0.47)	0.73 (0.44)
Evaluating Data	0.72	0.65

	(0.45)	(0.48)
Sharing Data	0.18 (0.38)	0.19 (0.40)
Storing Data	0.18 (0.39)	0.22 (0.41)
Sample	290	120

Notes: Data from undergraduate research posters archived in BYU ScholarsArchive for the FHSS Mentored Research Conference and Life Sciences Competition. Posters that used both primary and secondary data (N = 11) were excluded. T-tests were run to compare the outcomes between posters using primary data and posters using secondary data. * $p < .10$, ** $p < .05$, *** $p < .01$.

We also compared data literacy competencies by student experience, using their class standing as a proxy for experience. As students progress through their coursework and gain experience, we would expect them to grow in data literacy. Our findings, reported in Table 6, demonstrate that overall, upperclassmen demonstrate similar data literacy competency levels compared to less experienced students. The exceptions are in using mixed methods and describing data, which upperclassmen were more likely to do in their research compared to underclassmen. On the other hand, underclassmen were more likely to share data; all data sharing from students with class standing information came from underclassmen. These results should be interpreted with caution as the sample size for underclassmen is small, limiting the statistical power.

Table 6. Proportion of Articles with Data Literacy Competency Practice, by Class Standing

	Upperclassmen	Underclassmen
Manually Coded Measures		
Methods	0.84 (0.37)	0.74 (0.45)
Used quantitative method	0.77 (0.42)	0.81 (0.40)
Used qualitative method	0.11 (0.31)	0.07 (0.27)
Used mixed methods	0.13** (0.33)	0 (0)
Used primary data	0.85 (0.36)	0.78 (0.42)
Used secondary data	0.19 (0.39)	0.22 (0.42)
Describe data	0.31 (0.46)	0.26 (0.45)
Analyze data	0.79 (0.41)	0.81 (0.40)
Clean data	0.18 (0.39)	0.22 (0.42)

Evaluate data	0.08 (0.27)	0.04 (0.19)
Include form data citation	0.06 (0.24)	0.04 (0.19)
Share Data	0*** (0)	0.07 (0.27)
Store Data	0 (0)	0 (0)
Number of data tables	0.37 (0.77)	0.52 (0.98)
Number of data visualizations	3.73 (2.20)	3.89 (2.62)
Automated Indexing Measures		
Describing Data	0.75** (0.43)	0.56 (0.51)
Cleaning Data	0.43 (0.50)	0.44 (0.51)
Analyzing Data	0.67 (0.47)	0.59 (0.50)
Evaluating Data	0.72 (0.45)	0.74 (0.45)
Sharing Data	0.19 (0.39)	0.22 (0.42)
Storing Data	0.20 (0.40)	0.11 (0.32)
Sample	205	27

Notes: Data from undergraduate research posters archived in BYU ScholarsArchive for the FHSS Mentored Research Conference and Life Sciences Competition. Posters by students with unspecified class standing (N = 206) were excluded. T-tests to compare the outcomes between upperclassmen (juniors and seniors) and underclassmen (freshmen and sophomores). * p < .10, ** p < .05, *** p < .01.

Appendix C provides visualizations that show the changes in data literacy competencies demonstrated in the posters over time. Since the year ranges for the social science and life science posters differ and because of unique trends within each discipline, the effects over time were evaluated within each discipline rather than with all the posters together. Correlating the manually coded measures with the year led to very weak correlations (less than +/- 0.30). The graphs in Appendix C clearly show the lack of positive or negative trends over time. Most of the data literacy competencies we measured fluctuate over time but are within a fairly consistent range. Some large jumps exist at the beginning and end of the poster time periods, likely due to smaller sample sizes and greater variance in those years.

A few patterns stand out, contributing to our understanding of students' data literacy competencies applied to research. Looking at the trends in methods used in social science research, we see a decline in quantitative research and an increase in qualitative research. This mirrors trends in the social sciences throughout the second half of the twentieth century (Alasuutari 2010). It also demonstrates that students are being exposed to and using a greater variety of data and analysis methods. Another trend in the social sciences is a slight increase in data sharing, as measured through data sharing terms identified through automated indexing. This is a positive sign and mirrors trends in social science research, including data management and sharing mandates from major grant funders.

In the life sciences posters, there are some trends over time in the number of figures or data visualizations students use. The average number of figures used reached a high of 6.57 in 2019, steadily decreasing to about 3.5 by 2022. This supports anecdotal evidence from the poster conference organizers that student poster design has improved over time and suggests that students may be incorporating simplification, which is a best practice in research poster design (Rossi, Slattery, and Richter 2020; Siedlecki, 2017).

Discussion

How and what kind of data do undergraduates use in research?

While undergraduate students may still be learning about the research process and data skills, data plays a central role in most undergraduate research. Most students use data (89% of posters include a data practice or section from the manual coding; 92% of posters include terms related to data skills from the automated indexing). The majority report their data in some way, commonly in a results section (77%) and/or through tables or figures (96%). In both the social sciences and life sciences, quantitative research and data are the most common, with some growth over time in qualitative and mixed methods research for the social sciences. For individuals supporting social science data and research, this suggests that it is important to be prepared to help students with a wider variety of data types and sources.

Looking at the most used secondary sources in the social sciences, librarians can support undergraduate researchers by becoming familiar with datasets collected by research groups on their campus and methods for finding data in shared repositories. Additionally, our study of secondary data sources suggests that becoming familiar with government sources for data, such as the National Longitudinal Survey of Youth (NLSY) from the Bureau of Labor Statistics or the genetics databases made available by the National Center for Biotechnical Information, can prepare librarians to support researchers from social sciences and life sciences respectively. Similar government data sources for research outside the United States include the annual macro-economic database (AMECO) from the European Commission's Directorate General for Economic and Financial Affairs and the Canadian Alcohol and Drug Use Monitoring Survey (CADUMS) from Health Canada. As undergraduate researchers in both the social sciences and life sciences use secondary data, librarians should also be prepared to help students evaluate the quality of data sources. Zilinski, Sapp Nelson, and Van Epps (2014) provide a great framework for teaching data source evaluation skills.

What data literacy skills are present or missing in undergraduate research?

In both the social sciences and life sciences, students demonstrate strength in describing and analyzing data. These skills are central to gaining insights from data and reporting important findings to an audience. These are also skills commonly taught in statistics or research methods courses. Students' strength in these areas suggests that they can conduct research. They can be trusted to be involved and contribute to the process. They will especially benefit when given the opportunity to take ownership of their work, such as by creating and presenting a research poster.

The research posters also display gaps in student application of data skills, particularly in the areas of sharing and storing data. One reason for this could be limited space on a research poster. Generally, details on data management are not as central to explaining research findings as details on data analysis. Additionally, uncluttered research posters are easiest to read (Rossi, Slattery, and Richter, 2020; Siedlecki, 2017), and posters are not meant to include everything that might be in a research article, let alone a full data management plan. However, the lack of data sharing and storing mirrors trends in published research articles, where data sharing and management lag behind other data practices despite journal data sharing policies (Marchant, 2023). Although undergraduate students at non-R1 institutions do not often receive data management training, they can and should learn data management best practices to prepare to succeed in future opportunities (Blackwood, 2021).

Are there heterogeneous effects?

Additional strengths and gaps in data literacy skills were revealed when we compared posters by discipline, type of analysis, type of data, and experience. Students doing research in the social sciences more often included discussion of the quality of the data. One reason behind this difference may be that students in the social sciences also used secondary data more often than life sciences students and may have felt the need to justify their choice of data source. On the other hand, students doing research in the life sciences more often used terms relating to data cleaning practices, suggesting a greater understanding or value of the process of preparing data for analysis.

When it comes to the type of method used, students using quantitative methodology more frequently demonstrated data literacy skills in several areas. This was particularly significant (statistically and practically) in the areas of describing and analyzing data. Part of this difference may be due to bias toward quantitative methods in our poster coding structure, which is a limitation of the study. Future research should explore patterns in how students describe, analyze, and communicate data in qualitative research. Burress (2022) found no statistically significant difference in the number of data practices displayed by students using quantitative methods compared to other types of methods. When data from our poster sample is analyzed by looking at the number of data practices in each poster (rather than the proportion of posters displaying a specific practice), we find a statistically significant difference at the 95% level. However, the differences here are not practically significant at a difference of less than half a data practice (manually coded measures: quantitative = 1.68; all others = 1.23; automated indexing measures: quantitative = 2.99; all others = 2.65).

For students using primary data, we found that they were less likely to include details about describing, cleaning, and citing their data compared to students who used secondary data. This adds complexity to the findings in Burress (2022), where students using primary data displayed more data competencies. Our findings suggest unique skills come with using secondary data, including added needs to describe, clean, combine, and cite data sets. Additionally, when we look at the number of data practices displayed in each poster, those using secondary data also included a slightly higher amount (manually coded measures: primary data = 1.48; secondary data = 1.76). The divergence in our results may be due to different proxy evidence for identifying data literacy practices, such as looking for discussion about the actions of data cleaning rather than just the mention of a software name, which students may not reference.

While we would expect students to grow in data literacy competency as they progress toward graduation, we found little evidence supporting this idea. Posters created by upperclassmen did not differ significantly in the data skills displayed. The only statistically significant differences were in describing data, sharing data, and using mixed methods. Of the posters created by upperclassmen, 75% included terms related to data description in the automated indexing measure, compared to 56% of posters from underclassmen students. However, as the manually coded measure for describing data does not show a statistically significant difference, this could point to a limitation in the automated indexing measure. It is only able to identify the presence of terms and not the context in which they are used. Another potential limitation in the automated indexing measure is that it can only cover the most used terms for a data practice, but there are many ways a student may describe data. In a similar vein, 7% of underclassmen posters included data sharing according to the manually coded measure, while no upperclassmen posters included data sharing. Again, the paired measure, in this case automated indexing, was not statistically significant. Upperclassmen also more frequently used mixed methods, with 13% of upperclassmen posters using mixed methods and no underclassmen using mixed methods. This suggests that upperclassmen may be prepared for more nuanced research.

Overall, our findings suggest that more school experience does not significantly improve data skills. This analysis is limited by the small sample size for underclassmen, which restricts the power and precision of results. It is also important to note that this analysis did not compare the same students over time, and the sample of students doing research at any experience level is biased toward more achievement-oriented students. This may account for our finding that there is little difference by experience level.

Conclusion

Through our content analysis of student research posters, we found that students display both strengths and gaps in their data skills. What do these findings mean for students, librarians, and universities? An important implication is that students have the capacity to perform research. They can contribute significantly to research projects and have important skills with data, including describing and analyzing data. Giving students opportunities to take ownership of research may help them continue to increase their data competency and confidence. We advocate for more universities and faculty to provide opportunities for undergraduate research and encourage and mentor students throughout the process.

It is also clear that there are gaps in student skills. Across the board, students displayed little evidence of sharing or managing data. There is conflicting evidence of skills in evaluating and cleaning data. With the additional finding that greater experience in school is not related to greater data literacy skills, there is clear room to improve training in these core data skills. While statistics and methods courses appear to be doing a good job of preparing students to describe and analyze data, more emphasis can be placed on cleaning data in these courses. Condon, Exline, and Buckley (2023) recommend partnerships between librarians, instructors, and other campus support units to help students overcome the hurdle of learning technical skills.

Additionally, targeted training on skills relevant to managing and communicating about research data would be beneficial. This is a great place for librarians to provide additional support through workshops, one-on-one training, or online tutorials. One example is the great data visualization and infographic checklist created by Kapel and Schimdt (2021). Incentivizing data practices through updating poster rubrics or awards to specifically include data may also help students recognize the importance of data literacy and more clearly articulate their data practices.

Our sample of undergraduate research posters is limited to students from one university, so the findings are not fully representative of the broader global undergraduate population. Additionally, by conducting only a content analysis, our findings relate solely to the observed practices in the text and images of the research poster. Students may exhibit additional data skills that are not articulated in the research poster due to space or relevance constraints. Future research can build on this understanding of current data skills and practices in undergraduate research by evaluating student data skills in other contexts. More research is also needed to assess the impact of recommended interventions such as workshops, tutorials, or updated rubrics on improving students' data literacy.

References

Alasuutari, P. (2010) 'The Rise and Relevance of Qualitative Research', *International Journal of Social Research Methodology*, 13(2), 139–155, https://doi.org/10.1080/13645570902966056

Altintas, N. N. *et al.* (2014) 'The Use of Poster Projects as a Motivational and Learning Tool in Managerial Accounting Courses', *Journal of Education for Business,* 89(4), 196–201, https://doi.org/10.1080/08832323.2013.840553

American Association of Colleges and Universities (2024) 'High-Impact Practices', https://www.aacu.org/trending-topics/high-impact

Blackwood, E. (2021) 'Outside the R1: Equitable Data Management at the Undergraduate Level', *IASSIST Quarterly*, *45*(2), 1–12, https://doi.org/10.29173/iq1011

Burress, T. (2022) 'Data Literacy Practices of Students Conducting Undergraduate Research', *College & Research Libraries*, *83*(3), 434–451, https://doi.org/10.5860/crl.83.3.434

Brigham Young University (2022a) '*Annual Report of Sponsored Research*', https://spo.byu.edu/0000018e-c41e-dbed-afaf-e6ffceaf0002/2022-annual-report-final-pdf

Brigham Young University (2022b) '*Rank and Status Policy'*, https://policy.byu.edu/view/rank-and-status-policy

Brigham Young University (2022c) '*Rank and Status Professional Faculty Review Procedures'*, https://policy.byu.edu/view/rank-and-status-professional-faculty-review-procedures

Brigham Young University (2022d) '*Rank and Status Professorial Faculty Review Procedures'*, https://policy.byu.edu/view/rank-and-status-professorial-faculty-review-procedures

Calzada Prado, J., & Marzal, M. A. (2013) 'Incorporating Data Literacy into Information Literacy Programs: Core Competencies and Contents', *Libri*, *63*(2), 123–134, https://doi.org/10.1515/libri-2013–0010

Carlson, J., Fosmire, M., Miller, C. C., Nelson, M. S. (2011) Determining Data Information Literacy Needs: A Study of Students and Research Faculty. *portal: Libraries and the Academy, 11*(2), 629–657. https://doi.org/10.1353/pla.2011.0022

Carter, J. (2021) 'Developing a Future Pipeline of Applied Social Researchers through Experiential Learning: The Case of a Data Fellows Programme', *Statistical Journal of the IAOS*, *37*(3), 935–950, https://doi.org/10.3233/SJI-210844

Condon, P. B., Exline, E., & Buckley, L. A. (2023) 'Data Literacy in the Social Sciences: Findings from a Local Study on Teaching with Quantitative Data in Undergraduate Courses', *Evidence Based Library & Information Practice*, *18*(1), 61–75, https://doi.org/10.18438/eblip30138

Duckworth, J., & Halliwell, C. (2022) 'Evaluation of Higher-order Skills Development in an Asynchronous Online Poster Session for Final Year Science Undergraduates', *International Review of Research in Open and Distributed Learning*, *23*(3), 259–273, https://doi.org/10.19173/irrodl.v23i3.6238

Frost, M. E., Goates, M. C., & Nelson, G. M. (2023) 'The Benefits of Hosting a Poster Competition in an Academic Library', *College & Research Libraries*, *84*(4), 495–512, https://doi.org/10.5860/crl.84.4.495

Ghodoosi, B., West, T., Li, Q., Torrisi-Steele, G., & Dey, S. (2023) 'A Systematic Literature Review of Data Literacy Education', *Journal of Business & Finance Librarianship*, *28*(2), 112–127, https://doi.org/10.1080/08963568.2023.2171552

Gilmore, J., Vieyra, M., Timmerman, B., Feldon, D., & Maher, M. (2015) 'The Relationship Between Undergraduate Research Participation and Subsequent Research Performance of Early Career STEM Graduate Students', *Journal of Higher Education*, *86*(6), 834–863, https://doi.org/10.1080/00221546.2015.11777386

Howell, L. (2024, March 11) Personal conversation with C. Jeffrey Belliston.

Howell, L. (2024, August 17) Personal email communication with C. Jeffrey Belliston.

Kapel, S., & Schmidt, K. (2021) 'A Student-focused Checklist for Creating Infographics', *Reference Services Review*, *49*(3), 311–328, https://doi.org/10.1108/RSR-07–2021–0042

Kinikin, J., & Hench, K. (2012) 'Poster Presentations as an Assessment Tool in a Third/College Level Information Literacy Course: An Effective Method of Measuring Student Understanding of Library Research Skills', *Journal of Information Literacy*, *6*(2), 86-96, https://doi.org/10.11645/6.2.1698

Logan, J. L., Quiñones, R., & Sunderland, D. P. (2015) 'Poster Presentations: Turning a Lab of the Week into a Culminating Experience', *Journal of Chemical Education*, *92*(1), 96–101, https://doi.org/10.1021/ed400695x

Mabrouk, P. A. (2009) 'Survey Study Investigating the Significance of Conference Participation to Undergraduate Research Students', *Journal of Chemical Education*, *86*(11), 1335–1340, https://doi.org/10.1021/ed086p1335

Marchant, M. (2023) 'Teaching by Example: Evidence of Data Literacy Competencies and Practices in Top Economics Journal Articles', *Journal of eScience Librarianship*, *12*(3), https://doi.org/10.7191/jeslib.757

Pratoomchat, P., & Mahjabeen, R. (2023) ‘Building Research Skills through an Undergraduate Research Project on Local Community’, *Scholarship and Practice of Undergraduate Research*, *7*(1), 64–70, https://doi.org/10.18833/spur/7/1/10

Rossi, T., Slattery, F., & Richter, K. (2020) ‘The Evolution of the Scientific Poster: From Eye-sore to Eye-catcher’, *Medical Writing*, *29*(1), 36–40.

Ruth, A., Brewis, A., Beresford, M., & Stojanowski, C. M. (2023) ‘Research Supervisors and Undergraduate Students’ Perceived Gains from Undergraduate Research Experiences in the Social Sciences’, *International Journal of Inclusive Education*, 1-18, https://doi.org/10.1080/13603116.2023.2288642

Shorish, Y. (2015) ‘Data Information Literacy and Undergraduates: A Critical Competency’, *College & Undergraduate Libraries*, *22*(1), 97–106, https://doi.org/10.1080/10691316.2015.1001246

Siedlecki, S. L. (2017) ‘How to Create a Poster That Attracts an Audience’, *The American Journal of Nursing*, *117*(3), 48–54, https://doi.org/10.1097/01.naj.0000513287.29624.7e

Stark, E., Kintz, S., Pestorious, C., & Teriba, A. (2018) ‘Assessment for Learning: Using Programmatic Assessment Requirements as an Opportunity to Develop Information Literacy and Data Skills in Undergraduate Students’, *Assessment & Evaluation in Higher Education*, *43*(7), 1061–1068, https://doi.org/10.1080/02602938.2018.1432029

Stegemann, N., & Sutton-Brady, C. (2009) ‘Poster Sessions in Marketing Education: An Empirical Examination’, *Journal of Marketing Education*, *31*(3), 219–229, https://doi.org/10.1177/0273475309344998

Tanner, J. S. (2024) ‘*Envisioning BYU: Learning and Light*’, Brigham Young University: Provo, Utah.

Worthen, K. J. (2017, August 28) ‘*BYU: A Unique Kind of Education*’, https://speeches.byu.edu/talks/kevin-j-worthen/byu-unique-kind-education/

Zilinski, L. D., Sapp Nelson, M., & Van Epps, A. S. (2014) 'Developing Professional Skills in STEM Students: Data Information Literacy', *Issues in Science and Technology Librarianship*, (77), https://doi.org/10.29173/istl1608

Endnotes

[1] Maggie Marchant is the Economics, Finance, and Social Science Data Librarian at Brigham Young University. She can be reached by email: maggie_marchant@byu.edu.

[2] Jeff Belliston is the Associate University Librarian for Administrative Services at Brigham Young University. He can be reached by email: jeffrey_belliston@byu.edu.

[3] The earliest usage of this phrase that can be found on the BYU website is in an address by then BYU President Rex E. Lee on 27 August 1990 (https://speeches.byu.edu/talks/rex-e-lee/mt-everest-found-byu-undergraduate-education-can/). In his installation of and charge to the current president on 19 September 2023, Elder D. Todd Chistofferson, Chairman of the Executive Committee of the BYU Board of Trustees stated, 'Since Brigham Young University is first and foremost an undergraduate teaching institution, I charge you to elevate that core mission' (https://speeches.byu.edu/talks/d-todd-christofferson/installation-and-charge/).

[4] https://scholarsarchive.byu.edu/fhssconference_studentpub/

[5] https://fultonconference.byu.edu/attending-the-conference

[6] https://scholarsarchive.byu.edu/library_studentposters/

[7] https://guides.lib.byu.edu/2024postercomp/judging_criteria

[8] Truncation used to ensure that all forms of a word were included.