Investigating teaching practices in quantitative and computational Social Sciences: A case study

Rebecca Greer[1] and Renata G. Curty[2]

Abstract

Data education is gaining traction across disciplines and degree levels in higher education. Teaching data skills in the Social Sciences in today's data-driven world is vital for preparing the next generation of data-literate and critical social scientists. The ability to identify, assess, analyze, and communicate well and responsibly with data is key for scholars and professionals to navigate dynamic and expansive information ecosystems. This paradigm shift demands instructors to adapt their curricula and pedagogy to advance students' computational and statistical knowledge. This paper presents some of the findings from a local report of a larger national project which explored pedagogical techniques and instructional support needs for teaching undergraduates with quantitative data in the Social Sciences. Results revealed that the core learning goal of instructors is to develop students' critical thinking skills with data, including the conceptual understanding of the research methods employed in the field; the ability to critically evaluate research methodologies, findings, and data sets; and prowess using quantitative and computational tools and technologies. A recurring theme across interviews was students' fear of math and technology and the challenges these fears pose to data-related instruction. Instructors value participation in a community of practice and are eager for more institutional support to advance their computational skills. Based on these findings, we suggest avenues for academic libraries to further develop services, activities, and partnerships to aid data instruction efforts in the Social Sciences.

Keywords

Data Literacy; Statistical Literacy; Computational Literacy; Social Sciences; Data Pedagogy

Background and Motivation

A thriving number of initiatives have emerged in the last decades as a response to the growing need to equip students with foundational data skills to succeed in our data-driven world. Carmi et al. (2020) articulate that data literacy (DL) is critical to achieving data citizenship, combining data thinking, doing, and participation. *Data thinking* entails the critical understanding and evaluation of the data. *Data doing* refers to how individuals engage with data and manipulate it in a more practical way, whereas data *participation* refers to proactively implementing transformative approaches with data to anticipate problems, and form solutions. Libraries play an essential role in helping to prepare the next generation of data literate and critical citizens capable of navigating the data landscape and lifecycle, including the means to identify and assess data sources, extract insights and patterns from data, and communicate them effectively and responsibly. Libraries are at the intersection of information exchange and commonly aid in the efforts of researchers, whether they be emerging or established scholars, to engage data to construct new knowledge.

Henderson and Corry (2021) describe DL as a broader term that encapsulates multiple 'aspects of statistical and assessment literacy, pedagogical knowledge and data-driven decision-making under one umbrella'. For this paper, we concur with their views and define DL as one's ability to confidently and effectively engage with data. Meaning that one should be able to collect data, identify and select existing data sources, read and process data in various formats, derive meaningful insights and accurate conclusions from data, as well as report and share data deliverables in a trustworthy and ethical way. We

also integrated OECD's (2019) understanding that data literacy encompasses both the technical and social aspects of data, and may include more overarching activities related to quality data management such as data documentation, curation, and citation.

Risdale et al. (2015) discuss data literacy skills and competencies in higher education and academic research. The authors propose a set of core skills and competencies for DL characterized by five dimensions:

1. Conceptual framework: general knowledge and understanding of data, including use and application.
2. Data collection: skills and knowledge related to quality data discovery and collection from multiple educational sources.
3. Data management: skills related to data organization, preservation, manipulation, curation, and security.
4. Data evaluation: skills related to data analysis, presentation, interpretation, and decision-making.
5. Data application: knowledge and skills needed to share and cite data, evaluate decisions using data, and work with data ethically.

Through an environmental scan and analysis of current Educational Data Literacy Competence Frameworks (EDL-CFs) and courses, Papamitsiou et al. (2021) extended these five dimensions into seven data-related core competence pillars for DL: 1) data location, access, and collection; 2) data comprehension; 3) data interpretation and transformation; 4) data use, application, and act on; 5) data analysis; 6) data evaluation and, 7) data management. These pillars generally map with the more holistic view of the data lifecycle phases used more extensively by research and academic libraries to design services and instruction to support data-related work in academic settings.

Data instruction opens up a wide array of opportunities to innovate teaching and learning while preparing globally competitive graduates to deal with real-world-oriented tasks. Nonetheless, this more authentic and experiential educational environment demands strong partnerships in curriculum planning, continuous professional development efforts, and computational support and lab infrastructure. While the Science, Technology, Engineering, and Mathematics (STEM) fields have been notoriously at the forefront of data education, more recently, the Social Sciences have begun to entertain more quantitative and computational approaches to address pressing contemporary social issues.

Along these lines, Conte et al. (2012) emphasize the role of social scientists in exploring complex social systems through quantitative and computational lenses, given the growing integration of technology in our everyday lives and the massive volume of data generated from social interactions. This exploration involves multidisciplinary approaches to better understand the emergence of behavioral patterns in societies, their relationships with one another, and how individuals, groups, and communities interact with their environments both online and offline.

In addition to valuing the contextual and ethical aspects ingrained in research processes methodologies, there has been a growing trend in the Social Sciences to incorporate data literacy and more statistical knowledge and approaches to automate work with data, including data cleaning, processing, interpretation, and visualization to course curricula (Stephenson & Caravello, 2007). However, such pedagogical transformation does not come without some challenges. These transitions are heavily

dependent on disciplinary and departmental traditions, as well as existing institutional support to facilitate conditions instructors can rely on. Thus, we followed an exploratory approach to investigate teaching practices and the factors that may hinder or promote quantitative and computational Social Sciences instruction at the University of California, Santa Barbara (UCSB). Through our research, we seek to understand not only how the university has been responding to this trend in education and adapting its pedagogy, but, more importantly, to identify what kind of support students and faculty need from the Library and other campus partners to advance their data skills.

This paper presents the findings of our local study as a part of a larger national project with Ithaka S+R involving 19 other academic institutions. The study's goals were: 1) Explore pedagogical techniques and support needs in teaching undergraduates with data and 2) Provide actionable recommendations for stakeholders within and outside the library to inform new services, and practices to advance data instruction in the field. Based on our case study, we hope to showcase opportunities to refine and enhance library services, activities, and partnerships to support computational and quantitative data literacy skills in the Social Sciences that can be relevant to other academic institutions.

Research Methods

Our local study[3] was carried out in coordination with Ithaka S+R and followed the guidelines and research design defined for the national project 'Teaching undergraduates with quantitative data in the Social Sciences'. We followed a qualitative and exploratory approach to understand the current practices of faculty teaching with data. The identification and recruitment of potential participants took into account the selection criteria pre-established by Ithaka S+R: a) instructors of courses within the Social Sciences, considering the field as broadly defined, and making the best judgment in cases the discipline intersects with other fields; b) instructors who teach undergraduate courses or courses where most of the students are at the undergraduate level; c) instructors of any rank, including adjuncts and graduate students; as long as they were listed as instructors of record of the selected courses; d) instructors who teach courses where students engage with quantitative/computational data.

A total of 22 instructors were invited to the study, and 10 consented to participation. Interviews were conducted between September 2020 and January 2021 and followed a semi-structured interview guide with questions on how students are directed to obtain and engage with data in the course curricula in tandem with instructors' professional development, and training needs to teach with data. Due to COVID and the campus shutdown, all interviews were conducted remotely over Zoom and were audio-recorded for transcription purposes. Interviews produced approximately 12 hours of audio recording. De-identified transcripts and metadata are available through Dryad[4].

We performed coding on MAXQDA 2020 using a mixed-method approach through the combination of both deductive (top-down) and inductive (bottom-up) strategies. We started with an initial code tree that echoed the main topics present in the interview prompts. This initial coding scheme evolved and was refined as we engaged more closely with the data through iterative rounds of readings and review, which helped us to identify, tag, and rearrange themes that emerged from our conversations with faculty.

Findings and Discussion

Through our recruitment efforts, we were able to interview 10 individuals from six departments within the Social Sciences. Three of our interviewees were from Anthropology, two were from Sociology, two from Communication, and one interviewee each from Economics, Global Studies, and Psychology & Brain Sciences. Predominantly, these individuals held a professorship with three individuals identifying their

rank as assistant or associate. The remaining two interviewees self-identified as lecturers. Of the courses taught by these individuals, there was an even split between upper and lower-division courses, with the majority of them being large-enrollment classes that are accompanied by a lab often facilitated with the help of a teaching assistant. The analysis of instructors' narratives allowed us to organize findings into four main categories, as described in the following sections.

Expected student learning outcomes and ways students engage with data

The desire to develop critical thinking skills and advance students' data literacy was consistently expressed across interviews. Faculty views of critical thinking can be widely defined as the ability for students to actively, and whenever possible, autonomously respond to problem-solving situations. More importantly, instructors desire students to understand the necessary procedural steps of working with data while constructing meaning from the data thoroughly and accurately based on a specific question or problem.

Most faculty described that their classes are designed around statistical tests students perform when they actively apply learned concepts to assess statistical reports. The instructor's objective is for students to identify flaws in analyses and misleading findings. The following excerpt reflects the need for these skills and how that might manifest for students:

> *'What is that? What does that even mean?' [...] So, those are the critical thinking skills I want them to have to be able to assess right away, if you know, some representation of data that someone is putting out there is problematic. And usually, you can tell it's problematic, just, by the way, they have graphed it, there are ways to graph things to make the pattern look less clear and to make a pattern that's not there look like an actual pattern. (UCSB 1).*

For most interviewees, teaching students how to correctly interpret data and identify inaccurate statistical findings is key to preparing students to be critical consumers of data. '*You don't want to just hear something and then take it in without being critical, you need to be a critical consumer to understand if you should believe what you're being told'* (UCSB 6). Similarly, UCSB 9 highlights that students are required to analyze published findings, which leads '*into a critical discussion of what are the accurate statements you can make based on these data, then which statements misinterpreting correlation as causation'*, given that this is a common misconception in statistics.

Some instructors also described the importance of their courses as means to increase students' professional skills that align with the job market: '*I kind of go through some general examples like that, kind of hitting some of the careers that I know that our majors tend to gravitate towards'* (UCSB 6).

Learning Goals

Interviews surfaced three main learning goals that reflect instructors' experiences teaching with data in the Social Sciences.

Conceptual understanding refers to an integrated view of theories, methods, and concepts, their possible applications, interconnectedness, and scenarios where they can be applied. It also reflects students' abilities to articulate reasonable questions which can be answered statistically. For example, a student learning about group comparisons should be capable of understanding the underlying theory and basic principles behind the most widely used statistical techniques for that purpose (e.g., T-Test, F-test, ANOVA, MANOVA), their main differences and relationships, as well as the specific assumptions (e.g., sample size, distribution) they must consider.

Critical evaluation represents one's ability to holistically understand and make an informed assessment of the methods and approaches followed by others, understand the meaning behind the outputs, and evaluate the validity and reliability of the assumptions or conclusions based on data. This learning goal also includes one's ability to identify limitations of collected data and identify potential ethical concerns. Critical evaluators can form a plan of action based on their conceptual understanding of disciplinary knowledge in tandem with their ability to identify issues or gaps in the data to synthesize meaning. Following the same example above, a student with such skills should be able to evaluate if a given test meets the required assumptions and is correctly employed to analyze the data to answer a specific research question in a particular scenario while being capable of understanding the analytical outputs presented to them. Relatedly, critical evaluators would be able to target deviations from an original research question and decide which would be the most appropriate test to answer that question and produce meaningful and effective reports. It thus translates from a general hypothetical scenario to a real-world solution.

Working with data and/or tools comprises students' prowess to engage directly with data sets, identify and select existing data sources, gather, manage, and manipulate data, as well as operate (at least at a basic level) tools that can help them to automate analyses and create visualizations to convey meaning. This entails the application of concepts and evaluation to perform hands-on problem-solving beyond analytical reasoning. To satisfy this learning goal, students should confidently work with the dataset and perform their chosen statistical test using a tool such as R, Excel, Google Sheets, Stata, or SPSS to produce meaningful outputs.

As illustrated in Figure 1, these learning goals are complementary to each other and might play a more or less important role depending on the specificity of the course. Some faculty expressed that they dedicate most of their courses to explain concepts and basic statistical principles. Others focus more on the evaluation of published studies and statistical reports. In contrast, others emphasize more hands-on practice with tools, and some try to balance all of these learning goals simultaneously. These goals complement one another, the visual depicted below shows the amalgamation of instructor responses.

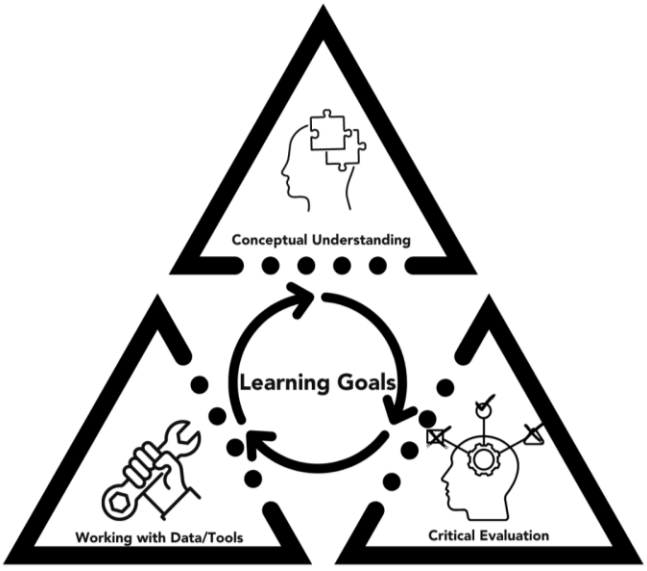

Figure 1 - Learning Goals

Expected learning goals were developed based on the clustering of skills interviewees desire to equip students with within their courses, as detailed in Table 1:

Table 1 - Expected Learning Goals and Skills Observed

Learning Goals	Skills	Definition 'Students should be able to…'
Conceptual Understanding	Develop hypotheses	Identify relevant questions that could be asked and answered with statistical data.
	Ground stats into the discipline	Articulate potential applications as well as the advantages of statistics in the context of their field.
	Master key statistical concepts	Identify variable types, units of analyses, and measurements
Critical Evaluation	Identify patterns	Spot and observe trends and correlations in data sets.
	Extract meaning from data	Read 'beyond the numbers' and extract relevant associations from the data.
	Marry concepts and procedures	Make informed decisions about the best approaches to explore and analyze the data.
	Correctly interpret outputs	Successfully evaluate and explain statistical analyses and their results.
	Write reports	Produce statistical reports and effectively communicate findings while following recommended styles and conventions.
Working with Data/Tools	Locate and access data	Search, identify and access available data sources.
	Basic coding	Feel more comfortable with tools that require some coding and writing basic scripts to automate statistical analyses.
	Perform basic analysis	Run statistical tests which are more common to their field.
	Create visualizations	Create meaningful graphical representations to represent findings.
	Use new tools	Know how to use different software and statistical packages that could help them to more easily and efficiently work with data.
	Test hypotheses	Perform hypothesis testing and verify possible correlations and relationships.

Evidence of Learning Goals in Instructional Praxis

The goal to establish conceptual understanding is described by UCSB 4 as they relay the importance of introducing students to the hypothesis development:

> *I try to show them the older hypotheses, the gaps in the hypotheses, and you know how those gaps are illustrated by certain examples from the collated data [...] that is to get them to think about, this is still a growing and developing field, perhaps they have some contributions to make and kind of trying to get them excited about it.*

Instructors strive for students to gain familiarity with potential applications of stats to answer research questions in their field while challenging them to exercise their logical reasoning. The following excerpt exemplifies attempts by an instructor to foster a welcoming environment for students to apprehend key statistical concepts while navigating the provenance and methods behind the data:

> *Whenever I introduce a variable, first of all, they have to understand, they're obliged to understand every aspect of the definition of the units in which it's measured and the real-world process by which somebody arrived at that number. [...] The question is, what units is it measured in, what does it capture, who made the number, who invented the number, what are the components of the calculation that were imputed, who imputed them (UCSB 8).*

Relatedly, UCSB 2 emphasized that their classes allow students to exercise their ability to 'think synthetically with some of the data and take in data sources from a bunch of different places that may not necessarily have obvious connections or sometimes have very obvious connections'.

The importance of students being capable of critically evaluating the data and effectively communicating their insights and inferences was expressed by some instructors who assign statistical reports as deliverables for their classes:

> *The other thing that I think is important is the writing process. […] They write a research paper, they introduce a problem, they review the literature on that problem [...] to address a gap, and then present their data and methods. And then, present their results, and then come back to the original question and say 'what does this tell us about this question?', or 'how does this help us move the literature forward?'. So I think that's a very useful skill to have and to be able to sort of communicating […] others what you've discovered through your analysis. (UCSB 4).*

UCSB 8 also sees the process of producing their own statistical reports as an opportunity for students to become more data literate and critical of other people's work:

> *I really want students to be able to not just calculate statistics and values but to understand what those values mean on the other side. [...] We give them a data set and they work their way through statistical analyses and then they report those analyses in APA style. [...] How to report those values to someone else because I think that helps them with actually reading empirical papers. So, when they practice writing a results section, I think they are better equipped to actually read the results section of the paper.*

The ability for students to search, find, and access relevant existing data sources was highlighted by a few faculty as a part of the skillset of their courses. UCSB 5, for example, described that they usually talk about 'how one might obtain data sets as well [and] some of the technicalities around [this process].' Such as, 'Finding data online using web scraper technology or using API technology to download data sets'. We observed, however, that instructors in this study most often supplied the data sets for class assignments, and only a few had students generate the data themselves. Instructors usually chose publicly available and previously de-identified data such as national statistics, datasets that do not involve human subjects, or even develop assignments around dummy data. To provide student access to the data sets, interviewees often relied on the institution's learning management system, GauchoSpace (Moodle).

Most faculty indicated that at least some part of their course workload covers some basic functionalities of the statistical packages and software, which can help students to compute statistics more easily. These demonstrations are usually provided during lab sessions, and in some cases, are complemented by step-by-step guidelines provided to students on how to use the tool. Microsoft Excel was the most commonly used tool by instructors, and in some cases, interviewees expressed the necessity to move from Microsoft Excel to Google Sheets to ease issues with access to paid software. Other software or tools referenced in the order of prevalence by interviewees were: R, SPSS, Stata, Q-GIS, and Eviews. Four of the 10 interviewees combined Excel with one of the other aforementioned programs for students to analyze or interpret data.

Beyond time in the classroom or lab, some interviewees described strategies for promoting extracurricular learning as means to help students to advance their data-related skills. UCSB 7, for example, created a series of optional short videos and demos that students can watch at their own pace while working on class assignments. UCSB 5 often recommends YouTube videos and Khan Academy courses, and UCSB 2 referred to their syllabus, which includes 'all kinds of external resources that are both associated with the university and wiki pages on the internet that are good for learning GIS and where people ask questions to learn how to do things.' A few faculty noted that internships and research projects with undergraduate students helped them become more well-versed in quantitative approaches and computational tools. However, they recognize that only a tiny percentage of students have participated in such activities.

Main challenges of teaching with data

Instructors' perceived challenges were mostly connected to learners' math and tech anxiety and helping them overcome these limitations without producing cognitive overload. Most interviewees acknowledged students' self-professed fears and obstacles students have had to overcome concerning their readiness to engage with statistics and affiliated tools to perform computations and produce outputs. Not only do instructors believe that students' entry knowledge is often limited, but they also recognize that the field still offers students few opportunities to develop statistical skills to gain confidence in using tools to automate their work, and that their classes are not able to fulfill all existing deficits.

In addition, the vast majority of the interviewees see Social Sciences at UCSB as less inclined to positivist traditions. Some interviewees expressed that their classes are the only opportunity for undergraduate students to interact more closely with quantitative data. The lack of options added to the fact that most interviewees signaled that students choose their majors based on their predisposition to soft sciences, which poses some challenges to instructors responsible for introductory courses on quantitative data.

> *I think the thing that's most relevant to classes involved in data analysis is just students' fear of math. And that's going to vary across the divisions, right? I mean, I'm sure engineering students have a lot less fear of math than Humanities and Social Sciences students do. (UCSB 6).*

UCSB 5 mentioned that students often make comments such as: 'What, it involves math? Logic? No, I don't want to do that.' Similarly, UCSB 9 mentioned that students are often quite 'upset about having to touch numbers and having to work with numbers in the first place', and later expounded:

> *[...] to actually teach them anything about how to do quantitative analysis in any serious way is very difficult, because they're not seeing any of it in any other classes, many of them are taking classes where they're actually actively discouraged from dealing with quantitative social science. They just don't have any training, many of them are math phobic. They have chosen [redacted,*

> *name of the course] because they are either afraid of mathematics or somehow, have some issue with it (UCSB 9).*

The fear of dealing with tech or the lack of digital dexterity, even among those courses which introduce basic software, such as Excel and Google Sheets, with no coding to perform statistical analysis, was also observed by some of the instructors. For example, one interviewee expressed the expectation that students should be more comfortable handling computer systems, but, unfortunately, that is not always true: 'And this is something which has sort of surprised me because I just assumed that over time, more and more people would be computer savvy, and it does not seem to be going that way.' (UCSB3). On the same note, UCSB 8 stated:

> *In fact, I've been quite surprised, for example even working with Excel some students have a difficult time and don't seem to have much experience in working with Excel or uploading something to R for example. I mean it requires a little bit of coding knowledge but the process is kind of similar to uploading a photo to an email. (UCSB 8).*

Besides the general understanding that most students in the Social Sciences are not as keen on learning statistical and computational approaches to analyze quantitative data, some faculty stressed the challenge of balancing statistical with computational instruction in one course. 'The whole point of the class is for them to learn how to use the software in addition to how to use the data that they use in the software' (UCSB2).

Some faculty expressed concern about the amount of time spent teaching students how to navigate and operate statistical tools and how that can take away their ability to focus on the lessons' content. This concern was presented as a justification for their choice to work with more basic computational tools. We believe that this may, at least partially, influence instructors to lower their expectations for students within their classes.

> *[…] we have them produce histograms, line graphs, you can do box plots, within SPSS, it has its own kind of data visualization. Nothing fancy. And the class, like I said, we keep the class very, very simple and this is not an advanced class. This is a class for students who have an absolutely terrible fear of math and a fear of quantitative data and so we keep it very simple. [...] Again, these are students who are not strong enough, resent having to take anything related to math or data analysis and I don't want to overwhelm them [...] (UCSB 6).*

Findings also show that most faculty have been adopting measures to mitigate such challenges, by minimizing obstacles with tools and employing ones that are easier to operate while working with vetted data sets where they have better control, but the data still demonstrates some of the basic statistical concepts and applications. While we understand that this approach seeks to accommodate students' fears, we believe it reduces opportunities for students to engage more critically with the data to apply learned concepts and tools to different contexts. On the one hand, having more control can help students better acclimate to statistics; on the other hand, it can narrow opportunities for students to practice their problem-solving and to translate general skills more autonomously to specific contexts.

Instructors' training and resource sharing

Instructional training and resource sharing varied among interviewees. Aside from their graduate education training, most faculty rely on professional development opportunities, such as academic

conferences and workshops, including Carpentry Workshops offered at UCSB Library to advance their teaching with data skills. The second most common method is through self-discovery, such as reading books and related literature on the topic or program, watching online video tutorials, and following trends with large technology companies such as Google LLC, Meta Inc. (Facebook Inc.), and Microsoft Corporations. Another less referred method observed was through inter- or cross-departmental collaborations with other faculty. In these few instances, colleagues were often consulted as a resource for independent research to fine-tune their techniques. In other cases, they consulted with colleagues for their expertise as an affiliated resource either for instruction or their research.

> *I've started to do some computational work or work that requires computational analyses with big data and I don't have the skill set to do that. I partner with computer scientists. So, I collaborate. I find people who have the data analysis skills that I don't. (UCSB 6).*

The majority of interviewees said they were either willing to share or have shared their instructional resources with students and/or fellow instructors. However, only half of the interviewees have used shared resources to develop course materials. One respondent, in particular, expressed that they would not be able to use shared resources or conversely share their own instructional materials due to the nature of their class being the only one of its kind in the department. Any resource sharing that could be done inter-departmentally would not be usable for their needs.

> *[...] you know that that's something I've entirely had to figure out on my own. And, you know, I get a lot of practice with that, because nobody else in my department teaches quantitative anything. So I've had to do a lot of it, and I just sort of learned through trial and error […] (UCSB9).*

Instructors described the benefits of engaging in a community of practice through which they can both reuse and share instructional materials. Yet, respondents often cited difficulty engaging in professional growth and training in this area due to a lack of time. Anecdotal events were referenced, such as a summer institute where instructors were prompted to reflect and retool pedagogical techniques as an opportunity to interrogate their teaching practice and reimagine their approach to measuring student learning. However, examples like these were limited as a robust, integrative opportunity to retool their instructional practices.

Types of support needed

The interactions we had with instructors confirmed our underlying assumption that there is a growing movement or desire toward more computational and quantitative-oriented Social Sciences. Some faculty acknowledged student attentiveness and interest when courses were oriented around digital topics.

> *Well, I think that there is a trend in the field that is sort of broadly one, where people are interested in anything digitally, I sort of attach the word digital to it, and people are all of a sudden, more attracted to it [...] because I think they're interested in developing translational skills. (UCSB 2).*

Conversely, most of the interviewees recognized that their departments are not ready yet to fully accommodate this trend, being unable to fulfill this demand since they are more heavily focused on qualitative research presently. During the interviews, most participants identified themselves as one of a few in their departments who conduct statistical research and engage in quantitative data-related

instruction. Some interviewees also emphasized how this qualitative orientation affects Social Sciences degrees at UCSB to advance this direction:

> *Faculty members are very, very few quantitative and that has been also a problem because we need more. In order to be able to be stronger in quantitative methods in [redacted - department name] here, and to be able to get more students that are quantitative, that want to do quantitative work, you need more faculty that does quantitative work. Because if not, you are signaling that this is a qualitative department. (UCSB 7).*

We also asked interviewees to describe the types of instructional support they need and receive to teach with quantitative data. As mentioned previously, most of the classes discussed were large enrollment classes with associated labs facilitated by teaching assistants. These teaching assistants were often referred to as invaluable to the course as they are the ones who align the critical intersections of theory and methodology with the available technologies.

> *In our courses they lead sections and so [...] in my classes that do data analysis they go over the homework problems with students and help them understand because students do data analysis by hand there and they lead the lab sessions. So, they are the ones teaching SPSS and showing them how to obtain the outputs to then interpret. (UCSB6).*

While these teaching assistants structure the practical use and application of the technologies for students, they also, at times, allow instructors of record to learn new data tools. For example, UCSB1 expressed that some students who have become teaching assistants were able to surpass the instructor's expertise with more advanced tools.

> *[...] they've sent me all of the material [...] all the books that they think are the best ones, they've sent me the YouTube videos that walk you through it. So I have gotten all these resources from two students who have previously taken my quantitative class [...]. So basically, I've taught them, they have gone beyond me, and now they're teaching me and that is the way learning should work […]*

It was not uncommon for instructors to express a desire for additional support with learning new programs and further support with technical aspects of the course, especially in a lab setting. For instance, interviewees expressed an interest in having a single, dedicated space on campus to host training on programs and technologies on teaching with data. This suggestion was often affiliated with comments where instructors expressed a lack of time or financial resources to pursue these interests independently. As such, they would value a structural intervention by their department, program, or the university as a whole to centralize professional development in this area.

In summary, our conversations with faculty surfaced the need for additional services and resources to support teaching with data in the Social Sciences on campus more broadly. The small-scale nature of this project, and the fact that not all relevant majors at UCSB were represented in our sample prevent us from generalizing results. We also acknowledge that drawing comparisons between pedagogical face-to-face approaches and instructors' strategies to accommodate classes to the virtual setting in the context of the pandemic was beyond the scope of this project, but could offer valuable insights for a future study. Despite these limitations, we believe our exploratory investigation of quantitative and computational data

teaching practices signals potential directions to address current challenges and limitations preventing data literacy from moving forward in the field of Social Sciences at UCSB.

Envisioning ways the library can leverage data instruction

Considering how academic libraries can actively intervene with the challenges experienced by instructors teaching with quantitative and computational data at UCSB, the co-authors surfaced an important distinction between those who teach credit-bearing courses and library programs and services. Instructors-of-record are bound by a curriculum limited by a quarter system which is 10 weeks in length. Library programs and services often follow the ebbs and flows of the quarter when responding to instructional and research needs. Yet, libraries are not bound in the same way by these conditions. Library personnel think expansively about how researchers at the institution engage with the continuum of the data lifecycle and necessary skills needed to work with data with these temporal conditions in mind. Library personnel must also remain accountable to the broad spectrum of academic skills and life experiences held by the various users they serve. UCSB is a Minority-Serving Institution and has an expansive portfolio of departments and programs that align with Social Science research practices. As expressed at the beginning of this paper, to engage in pedagogical transformation we must account for the disciplinary and departmental traditions while examining the institutional supports already in place, including our own. Our recommendations reflect possible avenues to support a variety of stakeholders to advance data instruction in the Social Sciences at UCSB, which could be influential for other academic libraries. In these recommendations, we recognize that this work will likely involve multidisciplinary approaches that can be disruptive to existing patterns in institutional and community contexts and learning modalities.

While library staff are contacted with practical research-based questions regularly, it is commonly noted that library staff need further support and opportunities to learn techniques when working with computational and quantitative data (Usova & Laws, 2021; Zaidane & Koizumi, 2019). Our findings did not directly surface a need for librarians or library staff to be involved in course curriculum development. It was commonly referenced that instructors lack sufficient time to engage in professional development activities to advance these skills. Hence, there is an opportunity for library staff to serve as consultants and deliver training to instructors and students at critical points of need. This requires staff to gain both practical experiences with using computational tools paired with a conceptual understanding of methodologies commonly applied in the Social Sciences if they are not familiar already.

Currently, UCSB Library provides a variety of information sessions and training on tools that can support statistical analysis. However, many of the interviewees were not familiar with these services. Library staff are encouraged to upskill and attend training in these areas, but there are no formal requirements nor structured communications to intentionally foster a broader community of practice in the library with statistical tools. Efforts can be made to improve communication and outreach not only to instructors but to groups, such as subject librarian liaisons, to express the utility of upskilling in these areas to support our academic community further. Active encouragement to engage in internal professional development can be led by distinctive groups, such as the Research Data Services (RDS) and the Interdisciplinary Research Collaboratory (IRC) at UCSB Library. These groups are best positioned to work with subject liaisons to identify current needs and tailor communications to departments when training or workshops are made available. A partnership such as this would provide novel opportunities for library staff to engage their departments and better apprehend how their liaison role can further support data literacy initiatives within the Social Sciences.

A common expression shared by many instructors was that they perceived students' fears often interfered with their ability to readily engage with the course content and that the limited time they have with students is usually not enough to overcome such fears and fulfill pre-existing knowledge gaps. Extracurricular activities could help mitigate these problems. There are many studies (e.g., Carlson, 2015) on ways to support successful extracurricular activities following the 'learning by doing' approach through hackathons, bootcamps, research projects, internships, and alike. As acknowledged by interviewees, students are still offered very limited opportunities to engage in similar activities, and the length of their course prevents them from planning these. At UCSB Library, some experiential and immersive learning activities around data are being considered for adoption through the newly launched RDS workshop series, with planned sessions where attendees will be invited to bring their own data to learn new skills, and have an opportunity to showcase their project deliverables.

In an effort to address potential issues of equity and access with immersive learning experiences with data literacy, the library is advancing projects to accommodate remote and asynchronous learning through modularized digital materials. Currently, a partnership is being established between RDS and Teaching & Learning (T&L) at UCSB Library to integrate these materials into course curricula. RDS holds expertise in the different stages of the research data lifecycle, including data gathering, cleaning/wrangling, analysis, documentation, archiving, and preservation. T&L holds expertise with foundational information literacy skills, pedagogy, instructional design, and educational technologies, and often provides instruction to lower-division undergraduate students. To help to allay student fears, RDS and T&L propose to create a foundational data literacy instructional module that can be readily deployed within the campus learning management system (LMS) and used asynchronously in credit-bearing courses. Pre- and post-assessments can be integrated with the module to measure student gains for both the instructor and the liaison librarians who may work with the class. This instructional delivery method has been successful when working with other entry-level courses, such as in the Writing Program at UCSB. Based on this prior experience, T&L and RDS will engage relevant stakeholders in the design process of this module to gather their input on the curricula and design elements, including intended learning outcomes and required assessment criteria. Once deployed, the asynchronous module can be paired with online discussion forums within the course LMS and drop-in office hours to connect students with library personnel who can further facilitate their research using quantitative data.

Further support for instructors can also be realized by modifying existing training and services at UCSB Library to better tailor to instructor needs. Carpentry workshops are offered continuously within the library's IRC in partnership with RDS and other campus volunteer instructors. Currently, these workshops are made available to all campus affiliates to gain facility with foundational data and computational skills. However, instructors noted that lack of time is a key barrier to participating in similar professional development activities. While working with liaison librarians to craft communications for those who teach computational skills in the Social Sciences, these messages could also incorporate surveys to gauge interest and availability to attend existing programming. A possible outcome from these efforts may be faculty-only workshops to promote and support the instructor's existing community of practice in a modality that best aligns with their scheduling needs.

Additional support for instructors may come in the form of train-the-trainer programming for Teaching Assistants and Associates (TAs). Many instructors rely on TAs to lead course or lab sections for large enrollment classes. Focusing on this audience to tailor services and programming, whether through hands-on synchronous workshops or the development of asynchronous course materials, may prove invaluable to a growing community of practice in the field and the next generation of instructors. In fact, engaging an audience of teachers who are students simultaneously can be key for the library to target

emerging needs that are not fully represented in the curricula as of yet. This aligns well with the library's interest in identifying and developing services and programming that support data literacy instruction generally, taking into account the whole data lifecycle. While it is imperative to pair our resources with curricular needs on campus, we are also positioned to advance the work of teachers in their dual roles as research practitioners who contribute to advancing disciplinary practices.

Data services and data education are logical and more recent outgrowths of the core role libraries have played for generations in educating the community they serve. While positioning the library as central to advancing data instruction in collaboration with departments, we also acknowledge the need to establish partnerships with other campus units and groups that could help us foster statistical literacy and data literacy more broadly at UCSB. Examples of potential collaborators include the DataLab coordinated by the Department of Statistics and Applied Probability, student organizations such as the Data Analysis and Coding (DANC) club, and interest groups such as the Quantitative Methods in the Social Sciences (QMSS) and the UCSB Research Data Community has members from various campus units to discuss data-related services, infrastructure, and instruction. Because libraries are often seen as hubs for central services on campus, we believe they may also act as natural facilitators to connect related but still siloed initiatives, while nurturing community building towards shared goals and efforts concerning data instruction. Mapping and engaging relevant campus units dedicated to advance data instruction to discuss and develop a more robust framework to support pedagogical approaches and learning assessments at the institutional level, in alignment with the whole data lifecycle continuum, is pivotal for preparing the next generation of professionals capable of exercising their data citizenship more confidently and profoundly.

References

Carlson, J., Nelson, M. S., Johnston, L. R., & Koshoffer, A. (2015). 'Developing data literacy programs: Working with faculty, graduate students and undergraduates', *Bulletin of the Association for Information Science and Technology*, 41(6), p14-17. Available at https://doi.org/10.1002/bult.2015.1720410608

Carmi, E., Yates, S. J., Lockley, E., & Pawluczuk, A. (2020). 'Data citizenship: Rethinking data literacy in the age of disinformation, misinformation, and malinformation', *Internet Policy Review*, 9(2), p1-22. Available at: http://hdl.handle.net/10419/218938

Conte et al. (2012). 'Manifesto of computational social science', *The European Physical Journal Special Topics*, 214(1), p325-346. Available at: https://doi.org/10.1140/epjst/e2012-01697-8

Henderson, J., & Corry, M. (2020). 'Data literacy training and use for educational professionals', *Journal of Research in Innovative Teaching & Learning*, 14(2), p232-244. Available at: https://doi.org/10.1108/JRIT-11-2019-0074

OECD (2019). Core foundations for 2030: conceptual learning frameworks. Available at: https://www.oecd.org/education/2030-project/teaching-and-learning/learning/core-foundations/Core_Foundations_for_2030_concept_note.pdf (Accessed: 12 Mar 2022).

Papamitsiou, Z., Filippakis, M. E., Poulou, M., Sampson, D., Ifenthaler, D., & Giannakos, M. (2021). 'Towards an educational data literacy framework: enhancing the profiles of instructional

designers and e-tutors of online and blended courses with new competences', *Smart Learning Environments*, 8(1), p1-26. Available at: https://doi.org/10.1186/s40561-021-00163-w

Ridsdale, C., Rothwell, J., Smit, M., Ali-Hassan, H., Bliemel, M., Irvine, D., ... & Wuetherick, B. (2015). Strategies and best practices for data literacy education: Knowledge synthesis report. Halifax, NS: Dalhousie University. Available at: http://www.mikesmit.com/wp-content/papercite-data/pdf/data_literacy.pdf (Accessed: 12 Mar 2022).

Stephenson, E., & Caravello, P. S. (2007). 'Incorporating data literacy into undergraduate information literacy programs in the social sciences: A pilot project', *Reference services review*, 35(4), p525-540. Available at: https://doi.org/10.1108/00907320710838354

Usova, T., & Laws, R. (2021).' Teaching a one-credit course on data literacy and data visualisation', *Journal of Information Literacy*, 15(1), p84-95.

Zaidane, A., & Koizumi, M. (2019). 'Roles of data librarians and research data services in academic libraries', Information and technology transforming lives: connection, interaction, innovation. Proceedings of the XXVII Bobcatsss Symposium, Osijek, Croatia, January 2019. Available at: http://bobcatsss2019.ffos.hr/docs/bobcatsss_proceedings.pdf (Accessed: 30 Mar 2022).

Endnotes

[1] Rebecca Greer is the Director of Teaching and Learning at the University of California, Santa Barbara Library. rrgreer@ucsb.edu.

[2] Renata Curty is the Social Science Research Facilitator at the University of California, Santa Barbara Library. rcurty@ucsb.edu.

[3] The study was IRB approved and was exempt by the UCSB's Office of Research in July 2020 (Protocol 1-20-0491).

[4] Curty, Renata G.; Greer, Rebecca; White, Torin (2021), Teaching undergraduates with quantitative data in the social sciences at University of California Santa Barbara, Dryad, Dataset, https://doi.org/10.25349/D9402J