Teaching with Data in the Social Sciences at Virginia Tech: An Ithaka S+R Local Report

September 2021

Liesl Baum, Center for Excellence in Teaching and Learning, Virginia Tech
Julia Feerrar, University Libraries, Virginia Tech
Kayla B. McNabb, University Libraries, Virginia Tech
Nathaniel Porter, University Libraries, Virginia Tech

Introduction

ITHAKA S+R Study

This report recounts an exploratory investigation of the needs and experiences of instructors who teach with data in the social sciences at Virginia Tech. It presents the findings and recommendations of one of 20 institutions taking part in this ITHAKA S+R project titled *Teaching with Data in the Social Sciences*. In the past, data literacy skills have been more commonly associated with disciplines in the hard sciences. However, encountering and working with data has become an integral part of every discipline as well as daily life. With this study, ITHAKA S+R and the research team hope to better understand and articulate teaching practices and support needs among instructors in the social sciences.

Virginia Tech Institutional Context

The curriculum at Virginia Tech is designed to meet the needs of students at all stages of their development. Many instructors are engaged in designing and teaching First-Year Experience (FYE) courses with the intention of supporting students through the transition from high school to college, broadly and into their major specifically. In addition, the general education curriculum, titled Pathways to General Education (formerly the Curriculum for Liberal Education), is designed as a “vibrant, flexible, and meaningful general education program, one that helps them to integrate their learning across disciplines and tackle challenges of the future” (Pathways to General Education, 2021). The faculty designing for both FYE and Pathways are responsible for incorporating the Essential Practices (FYE) and Core Concepts (Pathways) in addition to their core or disciplinary content. Several instructors who were interviewed for this project were, and remain, involved in Pathways and/or FYE.

In addition to the FYE and Pathways programs, part of the Virginia Tech strategic vision in recent years has centered on the establishment of Destination Areas (DAs), which are areas of research and teaching that are uniquely designed in response to the strengths of the Virginia Tech faculty and community. These areas of focus encourage the building of transdisciplinary communities to “address complex problems that impact the human condition” (Destination Areas, 2021). Several curriculum plans, minors, and even degree programs have been

developed through instructor involvement in the DAs. This report contains insight into some of the work resulting from the DA committee on *Data and Decisions* (Data and Decisions, 2021) with several courses developed and taught by some of the instructor participants. Those courses have been collaboratively designed and curated to also meet student general education requirements through the development of a Pathways Minor (Data and Decisions Minor, 2020).

Methods

To assess current practices and identify needs and opportunities in supporting undergraduate teaching with data in the social sciences at Virginia Tech, we conducted semi-structured qualitative interviews with 11 instructors in both academic and support units.[1] These instructors span the university and integrate the use or exploration of data in their undergraduate courses in a variety of ways. Our study is approved through Virginia Tech's Institutional Review Board. Teams at each participating university conducted their own interviews around a set of core questions and collected data relevant to their own institution to use for both local reports and a comprehensive report across all participating universities.

Semi-structured interviews ensure that each participant is asked the same set of core questions covering key topics of research interest while allowing flexibility to pursue follow-up questions that clarify details and explore unexpected responses (Given, 2008). Our interview schedule, included in Appendix C, focused on both learner and instructor experiences, including how learners are asked to find, evaluate, and work with data, and the support and practices instructors have found helpful or necessary.

Due to COVID-19-related restrictions, all of the interviews were conducted virtually using the Zoom conferencing platform through the Virginia Tech license. Interviews were recorded using Zoom and transcribed using an automated transcription tool before being cleaned and anonymized by the study team. Analysis was conducted in two stages. Initial collaborative coding identified four key themes (used as subheadings below): *Foundational Literacies for Working with Data, Experiential Learning and Applied Skills, Professional Development and Pedagogical Skills,* and *Ethics and Equity in Working with Data*. Each team member then used an inductive qualitative coding process in Taguette (Rampin, Rampin & DeMott 2021) to develop analysis in their section and contribute to discussion of overall findings and recommendations from the team. Anonymized transcripts will be archived online at the Qualitative Data Repository (https://qdr.syr.edu) to promote transparency and enable others to learn from and expand on our work.

There are a number of key limitations to the methods used to prepare this report. The limited number of participants means that entire departments are excluded, and their practices may differ from participants. Second, semi-structured interviews do not provide the same level of flexibility to allow participants to guide conversation beyond the boundaries of the interview

[1] These departments and programs included Food Science and Technology, Political Science, Human Development and Family Sciences, Sociology, Urban Affairs and Planning, Computational Modeling and Data Analytics, Statistics, and University Libraries.

schedule, and like the interviews themselves, coding was neither as rigorously deductive as theory-driven qualitative models nor as emergent as grounded theory. Finally, many individuals and experiences that could provide insight were excluded because of the criteria used to select participants and questions. The findings below thus provide an accurate but incomplete view of undergraduate teaching with data in the social sciences at Virginia Tech.

Participants

The Virginia Tech research team conducted a total of 11 interviews over the course of five months. These participants were among those identified as teaching with data in preliminary interviews with key informants at Virginia Tech and represented a diverse sampling of instructors that met the criteria for the project, *Teaching with Data in the Social Sciences*.

From this sampling, there were four out of eight undergraduate, degree-granting colleges at VT, including the College of Liberal Arts and Human Science, College of Science, College of Agriculture and Life Sciences, and the College of Architecture and Urban Studies. Within those colleges, eight unique departments were represented in addition to a broad sampling of instructor rank and position, including one Associate Dean and one Research Director (see summary in Appendix D).

Finally, while many instructors are involved in the FYE and general education curriculum (intro and methods courses, typically), a small portion of our sample is engaged in the design and implementation of advanced courses, such as senior capstone (sometimes labeled as senior design). These courses are designed to be a culmination of the student experience at Virginia Tech and help strengthen the connections across all of the knowledge that was gained during their undergraduate experience. In addition, these experiences often connect students with industry partners and key stakeholders in and across disciplines.

Findings

Among the interviews conducted for this project, several themes emerged surrounding participants' teaching practices, values, and support needs related to teaching with data. This report details four thematic areas identified in this study: *Foundational Literacies for Working with Data, Experiential Learning and Applied Skills, Professional Development and Pedagogical Skills,* and *Equity and Ethics in Working with Data.*

Foundational Literacies for Working with Data

One of the most pervasive themes in the interview data was the role of multiple digital literacies as foundational to teaching and learning with data. Almost all participants described common challenges for students related to computer literacy, information literacy, mathematical or computational literacy, and data literacy. Participants discussed a lack of these literacies as a potential barrier to working with data in their courses, noting that students bring a wide variety of understanding, experience, and attitudes. Where some students seemingly lack confidence in

their skills and have fears about engaging further, conversely others appear overconfident in their skills and disengage from foundational instruction.

Instructors discussed a range of knowledge, skills, and attitudes under the framing of literacy. In terms of computer literacy, participants discussed common student fears and confusion about "how computers work." One participant noted that "students generally do not know a lot about computers, and so they struggle with understanding how to set directories, how things are stored and saved on a computer." Practices involved in file management, including organization, storage, and naming conventions, came up frequently as challenges. Similarly, several participants identified information literacy challenges for students, including citation management, searching, and critically evaluating information and "fake news." Multiple participants also described challenges related to math or computational literacy such as the ability to calculate a percentage. However, one participant distinguished these kinds of skills from a more in-depth understanding of the math behind statistical analyses, which she identified as something her students do not need to understand. Finally, instructors discussed data literacy challenges around defining what counts as data, understanding the difference between a raw data set and an aggregated table, as well as a basic understanding of data privacy.

Some participants expressed surprise related to student prior knowledge and skills related to literacies. For example, one participant shared:

> I came in thinking: these are students who grew up in the Information Age who've been on computers since they were little kids. They should be expert Googlers and expert searchers. And I found that it is absolutely not the case. That was a bit shocking to me. So I had to go back and sort of teach some strategies for thinking about how to do that.

As exemplified here, multiple instructors identified a need to adjust their expectations and check assumptions related to student prior knowledge in these areas. This was particularly salient for instructors teaching introductory and general education courses that do not require prerequisites. Instructors of these kinds of courses shared a focus on building non-judgmental, inclusive learning environments. As one participant shared, "I try to structure the class so that there's something for everyone. I have zero prerequisites, so I don't judge people who have never opened Excel before." And similarly, another noted, "At times we view ours as a safe space where they might mess up and not quite understand what they're doing…" Providing learners with low-stakes, flexible opportunities to engage with data is key.

Acknowledging these needs, participants described bringing students "up to speed" or "to the same starting level" through a variety of approaches, including by making time during lecture or in-class activities, with guest instructors, through instructor or teaching assistant office hours, or through online learning resources such as modules and tutorials. Some participants mentioned working with subject liaison librarians and data consultants to integrate in-class activities and assignments that support students' foundational literacies and introductory data skills. Online learning options, in particular, came up as a way to supplement technical skill development or

introduce concepts to be further explored during class time. When asked about the kinds of training or assistance that would be most beneficial to instructors, one participant responded:

> I think that having easy-to-use online training modules and materials that students can go through that you get assigned for homework. And that way instead of spending class time saying, “this is how you do this method,” you say “having learned the methods in your homework, we can now start applying them for the types of analysis that we want to do.” So that would be really helpful.

In alignment with this need, some interviewees had already made use of online learning resources such as LinkedIn Learning (formerly Lynda.com) tutorials or a set of Advanced Research Skills modules developed by library faculty for undergraduate researchers (MacDonald, 2021). Additionally, one participant described recording his own tutorial videos to supplement online instruction during the COVID-19 pandemic. Participants expressed differing views, however, on the likelihood that students would engage with these kinds of resources: one participant expressed concern that students would not make use of video tutorials on their own time, while another identified this as something her students wanted more of and were already seeking on their own through YouTube.

While much of the interviewees’ discussion around literacies centered on challenges, they also emphasized their long term value for learners. Instructors discussed literacies as foundational to their course material, core to professional skills and career preparation, and even more so as key to engaged citizenship and lifelong learning. One participant discussed the importance of building a basic understanding of data scraping, noting that, “for me, the coding is an exposure to what citizens see or hear or read on a daily basis.” Other participants emphasized the importance of critical thinking skills, developing the ability to “read the newspaper and maybe not take things just at face value,” and becoming informed consumers and advocates for change. Instructors identified data and other digital literacies as key to their lifelong learning and engagement.

Experiential Learning and Applied Skills

The previous section demonstrates how digital literacy is a critical foundation for applied data and interpretive skills that support program outcomes, experiential learning and, ultimately, career goals for many learners. Some analytics-heavy programs embed sequential opportunities to develop and apply data skills. Most undergraduate teaching with data, however, takes place in either general courses with no prerequisites or singleton statistics/methods courses in otherwise packed programs. In neither case is there guarantee of existing knowledge and skills or follow-up opportunities to solidify and apply skills developed in the course, leaving many instructors struggling to prepare students for data activities that support learning outcomes and future courses or careers. In this section, we will first briefly discuss how instructors in data-heavy programs work to build skills, before turning to how those in the more common situation of isolated teaching with data adapt to support key learning outcomes within the limited scope of a single course.

There was widespread agreement among participants that undergraduates can learn to think analytically and work with data most effectively through extended (multi-semester) cumulative engagement in applied data analysis. The most obvious place this occurs is in the increasing number of analytics-focused major programs, such as Computational Modeling & Data Analytics and Smart & Sustainable Cities. Instructors in both programs, as well as Statistics, cited scaffolded activities, targeted to focus on a small number of learning outcomes at a time, as critical to successful teaching with data. Activities early in programs reduced cognitive burden by providing pre-cleaned data and easily-replicable code templates or programs with simplified interfaces. More than one respondent reported emphasizing strong foundations in scientific thinking and internalizing workflows at this stage, rather than beginning with technical skills, with the ultimate goal of deeper understanding and engagement.

An example of this is the QQQ model described by a statistics professor:

> The idea is you have qualitative forms of thoughts, quantitative forms of thought, and then back to qualitative forms of thought. And when we end with a final qualitative form of thought, that's when you're trying to specify explicitly insights that you're making from data, right? ... Throughout the entire class, a part of the QQQ process is thinking about implicit biases, both from the scientists to the consumer, as well as biases that play in terms of implications in the final qualitative component of critical thinking.

As students built skills and confidence, instructors found success gradually increasing the complexity, flexibility, and ambiguity of analytic activities. Common shifts include moving from menu-driven to code-driven software, moving from curated and pre-cleaned data to identifying and preparing learners' own data, and from pre-specified tasks to developing and testing original hypotheses. These programs can culminate in a capstone course where learners conduct analysis of data provided by an internal or external partner unit, such as a business or nonprofit, much like an internship/externship but with more guidance and supervision from professors in addition to the partner units.

Most programs, however, are not primarily oriented toward analytics and many require only one or two courses focused on methods, statistics, and data analysis. In addition to the challenges of establishing a foundation of data literacy before practicing applied analysis, this also limits the ability to gradually build complexity while reinforcing core skills. Our participants identified four different practices that help learners overcome these limitations: including data as a part of all discussions in non-methods courses, a simplified version of the scaffolded approach described above, optional minors or specializations, and curricular or co-curricular experiential learning options that support data skill development for motivated learners. Each is discussed in turn below.

Perhaps the most basic but overlooked of these is normalizing the inclusion of data across a wide range of courses, including substantive and theory-oriented courses to model data-driven thinking. A sociology professor described this approach:

> I'm very data-oriented personally in my own research and I bring that to bear, to discuss with students, but I don't have them do anything with it... I will tend to use data there because I'm kind of, I'm a data person and I want things to be data-based when I discuss them. And I try to have any undergraduate student understand that I'm not discussing feelings, attitudes or beliefs or much of that with them unless we're talking about measured beliefs and so on... But the most basic way I do it is by providing tables and charts for them to see that would talk about relationships between various variables and so on. And make sure that they see those in charts as data and not as something somebody just thinks about.

This professor was not even sure they had anything to contribute to this project because they didn't see their teaching as "teaching with data" in the traditional sense, but it is clear both from the comments on the need for data literacy and the approaches of other instructors that reading, interpreting, and thinking critically about how data is presented and whether it supports specific conclusions is a baseline condition for meaningful original analysis.

With that foundation, required methods courses in programs without heavy analytics components can more rapidly move to analysis and improve the likelihood that graduates will be comfortable interpreting and/or performing analysis in their areas, including critically evaluating quantitative analysis in published literature.

One specific challenge identified by those teaching this type of class is the availability of paid software for students. While more technical degree programs often train students in free and open-source code-based analytic software (such as R or Python), the learning curve is often steep to successfully integrate those into less methods-oriented programs. Commonly-used analytic software, including SPSS, Stata, and NVivo, can reduce the cognitive burden by providing menu-driven interfaces that guide learners toward the most common options, but they are only available in a limited number of student computer labs, if at all, unless students pay for personal licenses. While some paid analytic software such as JMP and ARCGIS are available to VT students at no cost, either through student programs of the manufacturer or campus-wide licenses, they may not match disciplinary standards that graduates will be expected to be familiar with, whether in industry or graduate programs. Data access is another challenge; many datasets are only available through expensive licensed subscriptions or otherwise difficult to access. Finding non-US data that was available through open source options or a similar option was identified as a particular challenge by multiple participants.

In addition to paid software and data, there is a complementary issue of familiarity with basic tools likely to be used in industry. At least one respondent required students to use spreadsheet software like Excel, rather than specialized statistical software, after finding that many students have never had other exposure to even basic spreadsheets. Participants identified this gap in skills as putting the students at a disadvantage on the job market because of the wide range of positions requiring them.

While the practices above take place primarily through core major courses, the remaining two practices provide alternative options for motivated students looking to build stronger data skills. The first is an increasing number of minor specialization options available to students. Some, like Research & Data Practices in the College of Agriculture and Life Sciences, are connected to a single unit and provide enhanced training in skills relevant to existing disciplines. Others, like Integrated Security Studies or Data & Decisions, are intentionally transdisciplinary and expose learners to approaches they might not see at all in their home discipline. Each of these minors provides structured, cumulative training in applied data analysis without the need to overhaul entire programs.

Formalized, governance-approved, minor programs are not the only options for motivated students to enhance data skills through experiential learning, however. Experiential learning for undergraduates is a key emphasis at Virginia Tech and our participants described multiple successful student-oriented models. A number of instructors were connected to supplemental programs such as Data Science for the Public Good, a summer program where selected students learn data skills and prepare reports or dashboards to assist governmental or non-profit agencies. Others supported cohorts and individual students participating in fellowships and other national or international workshops. Participants also reported supporting student-initiated and led efforts including clubs for sports analytics and foreign relations, as well as special events like hackathons.

One particularly innovative non-program-related model is Data Bridge, a data consulting group in the libraries consisting of both undergraduate and graduate students, supervised by faculty in the Data Services department. Prospective Data Bridge students begin by taking an independent study course with a cohort of other students where library faculty provide training in skills like data management, databases, and visualization, as well as in collaboration. Those who continue can work as student wage employees in the libraries, providing direct support for data-driven research projects in response to requests from VT researchers. Student consultants can come from any program and work with a variety of tools, depending on their interests and needs. Not only does this link outstanding real-world experience with regular supervision, feedback, and support, but it also enables the libraries to support more and larger projects than our faculty alone could manage.

In summary, all instructors have the opportunity to support teaching with data in the social sciences, but how they do so depends on the structure of their programs and the courses they personally deliver. All instructors can integrate critical discussion of summary data, regardless of course topic, providing foundations for critical and scientific data-driven thinking. Methods courses can use scaffolding and targeting learning outcomes to introduce key analytic tools and concepts without overwhelming students who may have less background in computing, statistics, or data analysis. Analytics-intensive major and minor programs provide structure and recognition for more extensive engagement to support more independent applied learning and projects, while co-curricular experiential learning provides motivated learners a variety of other opportunities to build skills and impact real-world projects, whether or not they are in computationally-intensive degree programs.

Professional Development and Pedagogical Skills

The instructors identified to participate in this research exhibit several patterns not only in approaches to teaching, but ways in which they prepare for their teaching. Of particular interest are the descriptions participants provided about their own professional development and the ways they have been prepared to teach with data. Most alarming is that the majority of these instructors revealed they have very little to no formal training on teaching with data. In fact, several of them are engaged in disciplines that, historically, have not naturally lent themselves to incorporating large quantities of data in the curriculum. When asked if they had received any type of training on teaching with data, one participant replied,

> No, I'm making it up as I go along. I mean, I talked to a lot of people about what—I've talked to people at [external university] and other places like when I was developing the course. I couldn't find a textbook. That was, you know, I couldn't find a textbook. And so I was really reaching out to a lot of different courses that I could find on the web and things like that that seem like it might be in this ballpark….So no, I did not have any formal training in how to use data.

As a result of this lack of professional development, many of these instructors feel as though they are on their own to identify resources and appropriate methodologies. To address this gap in their own preparation, they do a lot of searching across networks, including Google searches, exploring their discipline-based research conferences, reaching out to other colleagues, and even exploring cross-institutional connections in order to identify shared resources.

Not surprisingly, however, several participants discussed their interest in sharing resources across this same, or similar, network. As they explore for themselves, develop their own teaching and learning materials, and identify the impact of those resources on student learning, instructors recognize the value of sharing their materials with other instructors. However, they face the challenge of finding the most appropriate or most effective way to share as well as communicate within this organically-developed network. One participant states,

> So when I go find stuff, like I'm on people's Githubs, I'm pulling stuff from, from places like that. [And] I'm creating it myself. We create any presentation or lecture yourself just for any of the classes or for just [undergraduate research program] or modules for teaching in that. I just don't know where I would put it, necessarily. So what I have done for a lot of the tutorials is we do have a YouTube channel for the lab where we have videos that I use in classes and for training and stuff like that. We've got those and then we have an OSF page where I put all our wikis and tutorials and stuff there because at least that's [the] methods that we're using. And then people will know we use that in the research and then that's also how the students were trained to do a thing. So we do have the OSF because I'd just, everything we do as a group just, goes there. And so if they're looking for me for a thing, at least it's there.

For many of the faculty at our institution, professional development is viewed as an opportunity to not only improve one's teaching, but also remain in service to the university. They participate,

at will, with little to no extrinsic incentive (i.e. additional funding, resources, etc.). The general view is that as a result of them participating in high-quality professional development, instructors are contributing to the growth, strength, and character of the university with the added benefit of also exhibiting a vested interest in student learning and success. As mentioned, however, in the area of *teaching with data in the social sciences*, there is a major gap in what is offered for instructors in terms of professional development. During the interviews, several interviewees expressed their interest and motivation to participate in professional development programs designed specifically to support their needs in this area. One instructor stated, “I would love just like a tailored series, you know...maybe like an informal certificate or something like that I would totally sign-up for.”

In addition to programs specifically designed to prepare instructors who are teaching with data, there were additional resources that were identified which could support the outcomes of this professional development. Beyond strictly pedagogical strategies, participants identified that elements such as support staff (i.e. multiple instructors, teaching assistants or learning assistants, graders, etc.) as well as access to tools (i.e. campus licenses for software and assets) would make a significant difference to them. In addition, interviewees also mentioned having a network of guest speakers as well as more knowledge of campus support offices (e.g. university libraries) would help them fill in some of the deficiencies they were experiencing in their own training and preparation. More specifically, the described deficiencies in helping students clean the data, finding data, managing mixed methods work, as well as training in basic programming (for both instructors and students). With the added support, instructors felt they could be much more effective and provide a much stronger experience for their students.

Ethics and Equity in Working with Data

The previous themes point toward more foundational issues relating to the implications of both ethics and equity in teaching and learning with data, particularly in the social sciences. If students at Virginia Tech are going to be skilled and thoughtful contributors to their future contexts, then these issues must be addressed carefully but also thoroughly.

Turning to the question of ethics, interviewees highlighted concerns about learners reviewing, collecting, communicating, and making decisions based on data. As noted in the foundational literacy section, one interviewee referred to their concerns about teaching overconfident, computationally-informed students who disengage from the more basic lessons despite lacking “the reasoning or ethical components.” This poses a particular challenge for instructors who want to teach learners to review data ethically and attempt to provide them with the tools needed to “see inside the blackbox that is the process of getting data.” This kind of illumination and ethical reasoning surfaced in several of the interviews as the instructors discussed how they prompt students to ask questions about the data: where it came from, who collected it, who gets included, who gets left out, etc. While these may seem like extremely basic questions, interviewees cited versions of this kind of reflection to help learners consider how they might engage with data more ethically. Some explicitly had learners articulate the ethical considerations of their projects in a written reflection while others accomplished this through structured thought experiments and discussions in class.

One interviewee described an example they share with students where they ask them to consider the harms that might be caused by combining the lesser-represented racial groups in a survey as an ‘Other’ category, particularly when the implications of the research for a racial group is important to the research. This example also bridges into concerns about how data are communicated. Particularly among the interviewees who work with political science students or others who will impact policy decisions, there was significant conversation around the responsibility of the instructor to make sure students understand data, with one saying:

> We have a deep ethical, moral responsibility to think about how data is understood. Because we’re, we’re after all teaching students that again, many of whom, though not all, will go into policy and careers where the implications of the way that they understand data and use it are directly impactful on our lives.

These students will be making decisions based on data and about how to present or share data that could impact all of us. On directing students through this analysis process, one interviewee said, “The way I do it is for them to understand that that’s the way you analyze the issues. That’s my argument is you analyze issues this way. My refrain is, what do the data say?” Many of our interviewees take this task of training students to be ethical users and producers of data very seriously.

In addition to the ethical issues inherent in this work with data, equity issues present themselves in teaching and learning with data in several ways. Many of these are outlined or referenced above, but the project team feels that it is important to explicitly call out these broader equity concerns for both instructors and learners. Here, we will consider the inequity in education or training experience, support from instructors, support from programs, access to technology, and access to other support resources.

As noted in the foundational literacies section, undergraduate students come into these classes with a wide range of experiences. Some are confident in their technical abilities while others struggle with basic computer skills. This disparity prompts instructors to expect no prior knowledge or consider how they might be able to support students who need remediation to perform work at the level expected in the course. Furthermore, as noted in the professional development section, the lack of previous experience and training is not limited to students. Many instructors interviewed for this project noted that they did not receive training in teaching with data—or in some cases training for teaching at all—and hypothesized that a lack of training in pedagogy and/or data skills may be why some colleagues avoid more in depth teaching with data altogether. This discomfort or lack of preparation can perpetuate harmful ideas about some fields not being as ‘data-y’ as others. Individual instructors or whole programs may shy away from either the more technical work or the more nuanced social justice or equity concerns in an effort to stay with what is comfortable. One interviewee noted that some instructors use certain datasets specifically to “avoid baggage” and focus on skills instead of impacts, and programs have to decide if this avoidance of discomfort is resulting in a side-stepping the core purpose of teaching students data skills in the social sciences.

Additionally, many interviewees noted experiential learning opportunities, such as internships and extracurricular activities, as ways that students develop and exercise these critical skills, but it is important to note that there are external factors—such as economics—that may limit students' abilities to engage with these experiences, putting those students at a disadvantage to their peers. One participant stated:

> And then finally, internships are invaluable. So many students that have these skills might have acquired them through internships where they'd been doing similar work or at least that they've developed skills that we don't have as part of our major courses.

Relying on these kinds of experiences to provide training in lieu of other, more curriculum-integrated offerings potentially perpetuates equity issues that disproportionately impact learners from marginalized groups and first generation students.

Other externally-imposed limitations we saw in equitable access to working with data include access to software used to perform data analysis, which varied widely among interviewees. As noted in the experiential learning section, some selected inexpensive or free software while others chose more expensive software that was often seen as a standard in the field or profession. This cost impacts learners differently, with some getting access to software through the library or their student accounts (such as Google) while others pay one-time or subscription fees to access software to perform similar tasks for their courses or projects. Additionally, one interviewee cited issues with hardware as a significant equity stumbling block in his course saying:

> I had students the first year I taught this who are trying to do Python, load in pretty not huge datasets, but like a couple hundred thousand rows using a not very good tablet. It would just completely crash their machine, and they would get frustrated. They wouldn't even be able to play around with the data because they just couldn't open it.

This instructor devised a solution to offload some of the processing demand for his students, but all instructors in the social sciences may not be comfortable, willing, or able to pivot mid-stream to address these kinds of problems for their students, leading to inequitable support for learners.

Finally, access to additional resources, such as level-appropriate textbooks, supplemental educational materials, and campus services were mentioned by nearly every interviewee. Some resources, such as those available through LinkedIn Learning and the University Libraries, are available to all students, but access to other resources vary significantly. Textbooks offer a good example here; there are specialized texts to support developing data skills in political science while instructors from other disciplines described assembling learning resources from a wide range of sources to meet their students' needs or—in one instance—writing their own textbook to address the topic appropriately within their context. Furthermore, access to data sources that meet the needs of researchers was a recurring concern, with some interviewees noting that the library provides access to the journals, databases, and programs that allow researchers to

generate analyses and others lamenting the lack of easily available resources for their students' research areas.

These ethical and equity-related concerns are not limited to teaching with data in the social sciences, but it seems that those who work in the social sciences are particularly attuned to the implications of teaching practice without context or considering potential impacts.

Recommendations

This project has highlighted a number of opportunities to further support teaching with data in the social sciences across Virginia Tech. Participants discussed a variety of needs related to student competency, access and technology, program structures, and pedagogy. In some cases, there are existing programs and services already working to address these needs and they may be further communicated or extended to reach new audiences. In others, addressing these needs could include new interventions that focus on students, instructors, and institutional support, as well as an interplay between all of these. University Libraries, the Center for Excellence in Teaching and Learning (CETL), Technology-Enhanced Learning and Online Strategies (TLOS), and Undergraduate Academic Affairs are some of the units already supporting this kind of work and will be key in moving forward.

We offer the following recommendations and next steps.

Develop Targeted Communication About Literacy Programs: Campus units such as University Libraries already offer programs related to foundational digital literacies for students (Digital Literacy, 2020). In order to further leverage and communicate these programs, library faculty and other partners could more intentionally identify and connect with instructors who teach with data, particularly those involved in First-Year Experience and General Education. This could include:

- Adding data literacy examples to an existing teaching toolkit for First-Year Experiences instructors.
- Creating language about digital literacy workshops and data literacy for liaison librarians to share with their departments.
- Sharing resources during existing professional development for instructors involved with Pathways General Education.

Potential partners: University Libraries, Office of First-Year Experiences, Office of General Education

Next step: Share report findings with digital literacy librarians.

Expand Shared Online Learning Modules: Online learning modules and videos came up in multiple interviews as a way to differentiate instruction across learner needs and supplement or prepare for in-class activities. To expand these kinds of resources, University Libraries and TLOS could partner with academic departments to curate or adapt existing resources and

develop new ones. Modules could be developed in line with the needs of particular programs and then adapted for others, or could be developed by a cross-disciplinary group with the goal of making them broadly shareable through the Odyssey Learning Object Repository (About Odyssey, 2021). Internal grant funding has supported these kinds of projects in the past.

Potential partners: TLOS, University Libraries.

Next step: Identify potential partners for an internal Pathways grant proposal to create shared course content.

Explore Peer Learning and Consultation Models: Some participants felt that their students were more likely to seek help from peers or teaching assistants. As a way to expand this kind of support, academic departments or other academic support units could explore consultation approaches, especially those incorporating peer education. This could reflect something akin to a writing center model for centralized data literacy needs or build on course- or program-specific peer advisor or teaching assistant models with train-the-trainer education.

Potential partners: CETL, University Libraries, SAIG, Writing Center

Next step: Talk to leadership in other units with peer consultation (SAIG, Writing Center) about challenges and best practices.

Expand Professional Development: In this study, we identified multiple needs for instructor development around teaching with data in pedagogically sound ways. Support structures might include offering workshops, clinics, and consultations through existing professional development structures, such as the Professional Development Network managed by TLOS, which already incorporates offerings from CETL and from University Libraries. While it may make sense to complete a broader needs assessment and/or curriculum mapping process before creating this kind of support, some of the potential needs as identified in this study are:

- Understanding student foundational skills and differentiating or scaffolding instruction accordingly.
- Simplifying data projects and assignments to target primary learning outcomes related to data (to avoid expecting a single assignment or course to meet all data needs).
- Strategies for teaching key processes and skills for evaluating and thinking with data, both for non-data-intensive (e.g. sociology of race and crime) and data-focused work (e.g. QQQ, etc.)

Potential partners: CETL, Carpentries at VT, University Libraries, TLOS

Next step: Facilitated instructor discussion (organized by report team) on Teaching with Data on 12/9 through Professional Development Network.

Software, Hardware and Data: Applied data analysis is necessarily dependent on the availability of relevant, accessible, and adequate data to analyze, and hardware and software to perform analysis. Specific challenges to consider include:

- Ensuring all students, regardless of finances or personal circumstances, have equitable access to personal computers and/or public computer labs with sufficient power for data analysis. While tablets, chromebooks, and older lower-powered computers can suffice for most course needs, most data and analytic software requires Windows or (sometimes) Mac computers with at least typical low- to mid-range computing power. This also includes maintaining an easy-to-locate centralized list of software available to students through labs, campus licensing, and free or reduced-cost student programs.
- Ensuring free or low-cost access to paid or licensed data and analytic software that is both adequate to analytic tasks and widely used in relevant academic, governmental, nonprofit, or private-sector settings where program graduates are likely to be employed. Software identified specifically includes SPSS, Stata, and most of all coding and qualitative data analysis software (CAQDAS), for which there are no public labs or computers with any of the four major packages (NVivo, MaxQDA, atlas.ti, and Dedoose) and there are no equivalent free alternatives.
- Providing access to learners and guidance to instructors and learners for data sources appropriate to both the topics and methods for applied data analysis in the social sciences. International comparative and geopolitical data was identified as a relative weakness in our collections here, as well as awareness of paid data sources and support for discovery.

Potential partners: TLOS (lab & class technology and lists), University Libraries (data sources & training)

Next step: Contact TLOS about existing resources and partnership.

Build a Support Network or Community of Practice: Finally, we see an opportunity for the formation of a support network or formal community of practice related to teaching with data. This network might take shape both across Virginia Tech and more broadly among multiple institutions, the goal being to provide support for instructors, create and share course materials, and communicate about professional development opportunities.

Potential partners: CETL, Pathways, University Libraries, TLOS

Next step: Facilitated instructor discussion (organized by report team) on Teaching with Data on 12/9 through Professional Development Network.

Conclusion

The landscape of teaching with data in the social sciences at Virginia Tech is complex, involving instructors with a variety of interests and expertise, courses across the undergraduate curriculum, and students with a range of prior knowledge and personal goals related to their

own learning. Some of the major challenges for teaching with data identified in this project include addressing the following:

- variety in student foundational literacies,
- ensuring equitable access to necessary software and hardware,
- actively engaging students in experiential learning with data,
- addressing complex conversations about data in context,
- and broadly ensuring that instructors have the preparation and development they need to teach effectively with data.

Many campus units are already working to address some of these challenges and are well-poised for further collaborative support. By building on existing digital literacies programming, expanding shared resources like online modules, and exploring models for peer education and further professional development, we can better prepare Virginia Tech students to learn, create, and take action with data throughout their lives.

References

About Odyssey. (2021). Retrieved September 20, 2021, from https://perma.cc/DA2Y-XVKL

Ayres, Lioness. (2008). “Semi-Structured Interviews” in *The SAGE encyclopedia of qualitative research methods*. Ed. L. M. Given. Thousand Oaks, CA: SAGE Publications, Inc. Pages 810-811. https://doi.org/10.4135/9781412963909.n420.

Data and Decisions. (2021). Retrieved September 29, 2021 from https://perma.cc/DPJ5-7M6E

Data and Decisions Minor. (2020). Retrieved September 29, 2021 from https://perma.cc/K4BN-M459

Destination Areas: Building Transdisciplinary Communities. (2021). Retrieved September 30, 2021 from https://perma.cc/T2AY-ANHH

Digital Literacy. (2020). Retrieved July 7, 2021, from https://perma.cc/5NNL-LRA5

Pathways to General Education. (2021). Retrieved September 29, 2021, from https://perma.cc/WS3S-H87Y

MacDonald, A. (2021). *Research Guides: Advanced Research Skills Program Spring 2021* Retrieved July 7, 2021, from https://perma.cc/SB6T-ZS9Q

Rémi Rampin, Vicky Rampin, & Sarah DeMott. (2021, July 17). *Taguette (Version 1.0.0)*. Zenodo. https://doi.org/10.5281/zenodo.5111814.

Appendix A

Project Timeline

The ITHAKA S+R research on *Teaching with Data in the Social Sciences* was conducted over the course of several months. The primary portion of the work took place between September, 2020 and March, 2021. These tasks included the curation of an interviewee list based on criteria that aligned with the objective of the project. Once the interviewee list was created, the research team contacted the first round of interviewees to schedule virtual interviews. Upon completion of the first round, a few additional interviewees were contacted to help the research team reach their goal of 12-15 interviews. In total, the research team conducted 11 interviews (see summary in Appendix D). It is important to note that all of the work on this project, including the interviews, were conducted virtually in order to remain in compliance with COVID-19 health and safety protocols.

Date	**Task**
May, 2020	Project Launch
June, 2020	IRB approval (determined to be not human-subjects research)
July, 2020	Interviewee selection
September, 2020	Round 1: Interviewee recruitment
October-December, 2020	Interviews (all interviews were conducted virtually due to COVID restrictions
January, 2021	Round 2: Interviewee recruitment
January-February, 2021	Interviews (all interviews were conducted virtually due to COVID restrictions
March-April, 2021	Transcription, coding, and analysis
May-July, 2021	Final report
October, 2021	Report published

Appendix B

Recruitment Email Script

Hello, _____:

I am part of the Virginia Tech team taking part in the Ithaka S+R research project exploring how our university supports teaching with data in the social sciences to better understand what is currently happening and ideally improve its support for this work in the future. We have identified you as a potential interviewee for this research due to your work with data in your courses. You can read more about our project in this blog post, and we're happy to answer any additional questions you might have.

We are hoping to meet with each of our interviewees for a roughly 1-hour interview during Fall or Winter 2020. If you would be interested in talking to us, please use the link below to schedule a time that works for you. If you do not find any times that work with your schedule, please let us know, and we will reach out to find a time.

[Interview Appointment Calendar]

Thank you for your consideration,
Liesl, Julia, Kayla, and Nathaniel

Appendix C

Interview Guide

Note regarding COVID-19 disruption I want to start by acknowledging that teaching and learning has been significantly disrupted in the past year due to the coronavirus pandemic. For any of the questions I'm about to ask, please feel free to answer with reference to your normal teaching practices, your teaching practices as adapted for the crisis situation, or both.

Background

Briefly describe your experience teaching undergraduates.

- How does your teaching relate to your current or past research?
- In which of the courses that you teach do students work with data?

Getting Data

In your course(s), do your students collect or generate datasets, search for and select pre-existing datasets to work with, or work with datasets that you provide to them?

If students collect or generate datasets themselves Describe the process students go through to collect or generate datasets in your course(s).

- Do you face any challenges relating to students' abilities to find or create datasets?

If students search for pre-existing datasets themselves Describe the process students go through to locate and select datasets.

- Do you provide instruction to students in how to find and/or select appropriate datasets to work with?
- Do you face any challenges relating to students' abilities to find and/or select appropriate datasets?

If students work with datasets the instructor provides Describe the process students go through to access the datasets you provide. *Examples: link through LMS, instructions for downloading from database*

- How do you find and obtain datasets to use in teaching?
- Do you face any challenges in finding or obtaining datasets for teaching?

Working with Data

How do students manipulate, analyze, or interpret data in your course(s)?

- What tools or software do your students use? *Examples: Excel, online platforms, analysis/visualization/statistics software*
- What prior knowledge of tools or software do you expect students to enter your class with, and what do you teach them explicitly?

- » To what extent are the tools or software students use to work with data pedagogically important?
- » Do you face any challenges relating to students' abilities to work with data?

How do the ways in which you teach with data relate to goals for student learning in your discipline?

- » Do you teach your students to think critically about the sources and uses of data they encounter in everyday life?
- » Do you teach your students specific data skills that will prepare them for future careers?
- » Have you observed any policies or cultural changes at your institution that influence the ways in which you teach with data?

Do instructors in your field face any ethical challenges in teaching with data?

- » To what extent are these challenges pedagogically important to you?

Training and Support

In your course(s), does anyone other than you provide instruction or support for your students in obtaining or working with data? *Examples: co-instructor, librarian, teaching assistant, drop-in sessions*

- » How does their instruction or support relate to the rest of the course?
- » Do you communicate with them about the instruction or support they are providing? If so, how?

To your knowledge, are there any ways in which your students are learning to work with data outside their formal coursework? *Examples: online tutorials, internships, peers*

- » Do you expect or encourage this kind of extracurricular learning? Why or why not?

Have you received training in teaching with data other than your graduate degree? *Examples: workshops, technical support, help from peers*

- » What factors have influenced your decision to receive/not to receive training or assistance?
- » Do you use any datasets, assignment plans, syllabi, or other instructional resources that you received from others? Do you make your own resources available to others?

Considering evolving trends in your field, what types of training or assistance would be most beneficial to instructors in teaching with data?

Wrapping Up

Is there anything else from your experiences or perspectives as an instructor, or on the topic of teaching with data more broadly, that I should know?

Appendix D

Summary of Participants

Total number of participants: 11

Unique colleges represented: 4

- College of Liberal Arts and Human Sciences (4)
- College of Science (3)
- College of Agriculture and Life Sciences (2)
- College of Architecture and Urban Studies (2)

Unique departments represented: 9

- Biochemistry
- Food Science and Technology
- Human Development and Family Science
- Mathematics
- Political Science (2)
- Sociology
- Statistics (2)
- University Libraries
- Urban Affairs and Planning (2)

Rank

- Professor (2)
- Associate Professor
- Assistant Professor (5)
- Collegiate Assistant Professor
- Associate Dean/Dean
- Research Professor/Director

Appendix E

IRB Approval Document - Not Human Subjects Research

VIRGINIA TECH

Division of Scholarly Integrity and Research Compliance
Institutional Review Board
North End Center, Suite 4120 (MC 0497)
300 Turner Street NW
Blacksburg, Virginia 24061
540/231-3732
irb@vt.edu
http://www.research.vt.edu/sirc/hrpp

MEMORANDUM

DATE: June 4, 2020

TO: Nathaniel Porter, Kayla McNabb, Liesl M Baum Walker, Julia Feerrar

FROM: Virginia Tech Institutional Review Board (FWA00000572, expires October 29, 2024)

PROTOCOL TITLE: Teaching with Data in the Social Sciences

IRB NUMBER: **20-482**

Based on the submitted project description and items listed in the Special Instructions section found on Page 2, the Virginia Tech Human Research Protection Program (HRPP) has determined that the proposed activity is not research involving human subjects as defined by HHS and FDA regulations.

Further review and approval by the Virginia Tech Human Research Protection Program (HRPP) is not required because this is not human research. This determination applies only to the activities described in the submitted project description and does not apply should any changes be made. If changes are made you must immediately submit an Amendment to the HRPP for a new determination. Your amendment must include a description of the changes and you must upload all revised documents. At that time, the HRPP will review the submission activities to confirm the original "Not Human Subjects Research" decision or to advise if a new application must be made.

If there are additional undisclosed components that you feel merit a change in this initial determination, please contact our office for a consultation.

Please be aware that receiving a "Not Human Subjects Research" Determination is not the same as IRB review and approval of the activity. You are NOT to use IRB consent forms or templates for these activities. If you have any questions, please contact the Virginia Tech HRPP office at 540-231-3732 or irb@vt.edu.

PROTOCOL INFORMATION:

Determined As:	**Not Human Subjects Research**
Protocol Determination Date:	**June 4, 2020**

ASSOCIATED FUNDING:

The table on the following page indicates whether grant proposals are related to this protocol, and which of the listed proposals, if any, have been compared to this protocol, if required.