

10-2021

Ithaka S+R: Teaching with Data in the Social Sciences

Gayle Schaub

Grand Valley State University, schaubg@gvsu.edu

Samantha Minnis

Grand Valley State University, minniss@gvsu.edu

Follow this and additional works at: https://scholarworks.gvsu.edu/library_sp



Part of the [Higher Education Commons](#), and the [Library and Information Science Commons](#)

ScholarWorks Citation

Schaub, Gayle and Minnis, Samantha, "Ithaka S+R: Teaching with Data in the Social Sciences" (2021).
Scholarly Papers and Articles. 69.

https://scholarworks.gvsu.edu/library_sp/69

This Article is brought to you for free and open access by the University Libraries at ScholarWorks@GVSU. It has been accepted for inclusion in Scholarly Papers and Articles by an authorized administrator of ScholarWorks@GVSU. For more information, please contact scholarworks@gvsu.edu.

ITHAKA S+R: TEACHING WITH DATA IN THE SOCIAL SCIENCES



October 2021

Full Report



Gayle Schaub, Samantha Minnis
Grand Valley State University

Introduction

This study is connected to a larger group of studies conducted simultaneously at twenty higher education institutions in collaboration with Ithaka S+R. Ithaka S+R is a not-for-profit research and consulting organization that helps the academic, cultural, and publishing communities, and has been hired by Grand Valley State University Libraries to coordinate this parallel effort and to provide guidance on research methodology and data analysis. Ithaka provided training and support materials for the researchers including questions for interviews and training in interviewing and coding. The anonymized aggregated data and analysis from this study will be used towards a comprehensive report written and made publicly available by Ithaka S+R.

Grand Valley State University (GVSU) is a comprehensive state-supported university. It has a main campus located in Allendale, Michigan, with four separate, additional sites in Grand Rapids and regional centers in Muskegon, Holland, Traverse City, and Detroit, Michigan. At the time of this report, the student body numbered 20,265 undergraduate and 2,834 graduate students. During the 2019-2020 academic year, 245 faculty members taught social science courses. As part of the university's general education program, students are required to take two courses from two different disciplines that introduce them to the content and methods of a social or behavioral science field. Because of these social science requirements and because of the importance of using data and the increasing importance of data literacy in a time when data is more available, influential, accessible, and, at the same time, more misunderstood and misused than ever, we felt it was important to participate in this study to learn more about the uses of data at GVSU specifically and to contribute our findings to the larger project.

Methodology

Participants were recruited through email. Potential participants were identified with the assistance of the GVSU office of Institutional Analysis as teaching in one of the following disciplines: archaeology, anthropology, economics, geography, history, legal studies, linguistics, political science/international affairs, Psychology, or Sociology. An email explaining the project and asking for participation was sent in batches (see Appendix A), resulting in eight responses. After a few weeks, a follow-up email was sent to those on the list who had not replied to the original email, resulting in four more responses.

Interested participants were invited to a one-hour meeting with one or both of the researchers. The interviews were conducted over Zoom, in compliance with campus and state restrictions and health concerns due to the COVID-19 pandemic. Twelve semi-structured interviews were conducted using questions generated by Ithaka S + R, focused on how instructors teach with data, the methods they use, the challenges they encounter, and their pedagogical rationale. The

interviews were recorded on the researchers' computers. Captions were generated using a secure video content management software account, edited into transcriptions and anonymized, and the videos were destroyed. One interview was lost due to a corrupted file.

To analyze the transcripts, both researchers read all transcripts to become familiar with responses. The researchers then visualized the information by writing the answers to the questions on a whiteboard. This process allowed the researchers to recognize patterns in the answers about uses of data, challenges with data, and support needs. The whiteboards were photographed (see Appendix B for an example) and then the themes were compiled into spreadsheets, one for each transcript, which were then collated into one master spreadsheet of themes.

Findings

Data Use: Teaching, Techniques, and Tools

Data and statistics are used widely in social science courses, in a variety of ways, to achieve a variety of pedagogical goals. From introductory level courses in anthropology, geography, sociology, economics, and psychology to advanced research methods and capstone courses in the disciplines, students locate, access, and collect data and use it to answer questions, identify patterns, explain social issues, visualize phenomena, forecast trends, make conclusions, and generate hypotheses.

As students engage in the many methods of data collection, access, and analysis, they are, as part of the learning process, introduced to ethical standards and practices in using data. One professor has her advanced economics students work with local non-profits, doing economic data analysis for a homeless shelter and other small organizations. Psychology research methods courses use personal interviews. Students in sociology courses work with community partners, using their data to identify patterns. In all of these contexts, students are taught about the importance of accuracy and confidentiality. The faculty interviewed discussed how they emphasize critical thinking about the data they're using: where it originated, where it's stored, or how it's destroyed, how they'll use it, and issues surrounding anonymity and subject appropriateness. While there is not yet one specific course in data ethics, most conversations touched on the ethical use of data at least generally, very often specifically, with concrete examples. Interviewees weave issues of data ethics into their course material and find ample opportunity to teach through facilitating opportunities for their students to engage in hands-on activities involving data.

In most cases, types of data and their uses are strongly dependent on course level. In introductory courses and general survey courses, particularly those made up of a majority of non-majors, data is primarily instructor-supplied, or students receive explicit instruction on how to access particular data sets from pre-selected websites or curated collections of data. Many of the courses make frequent use of publicly available government site content, such as the US Census, Centers

for Disease Control and Prevention (CDC), Bureau of Labor, Housing and Urban Development (HUD), US Geological Survey (USGS), and other national and municipal sites from around the United States. Others use curated collections from databases like Social Explorer, Statista, or The Digital Archeology Record (t.dar). Many professors create data sets designed to achieve a specific disciplinary learning objective. A psychology professor teaching intermediate and advanced research methods in psychology creates tables of “raw” data for students to use in undertaking factorial design. A professor of archeology mocks up dig sites for students to use in performing a nearest neighbor analysis. In both situations, the purpose of the exercises is not the collection, location, access, or validity of the data, it is the analysis, so the educators elect to use “teaching” data rather than real raw or student-collected data so that their students can engage more intently in the process of data interpretation and/or analysis. They can work toward understanding what it is that researchers do with the data they collect, and how that knowledge creation becomes scholarship.

Other innovative uses of data in the social sciences involve personal interviews, cemetery headstone surveys, global movement investigations, demographic studies, and death inventories. Archeology students collect headstone data from local cemeteries and make conclusions about the history of a based on the characteristics of that data. With their professor, anthropology students compile a “death inventory,” and together, examine their personal experiences with death. They analyze the results based on course readings, ideas, and discussions. Beginning economics students analyze the data of Fortune 500 companies, the demographic makeup of their boards, and consider how gender differences impact profitability. In an upper level sociology course, students work in groups to build a website documenting a global social movement, they locate data to tell the story of an environmental or social justice problem, and use data to support their proposed movement. Through sampling, affinity diagramming, histogramming, forecasting, factorial design, via interviews, field work, maps, websites, and more, students are searching for, collecting, compiling, analyzing, and interpreting data in myriad ways, throughout their social science studies.

Finding data is a process that consumes more time than many professors are able to afford. Many use data sets they’ve held onto for semesters. Others collaborate with departmental colleagues. Some take the extra step to collaborate with librarians to find usable data or, better yet, work with their students to model the search process and help them become comfortable, proficient locators of data. Because there is no one comprehensive source of data in any specific discipline, professors and librarians spend a good amount of time searching the Web, talking to colleagues, building collections, crafting library subject guides, lists, and activities designed to familiarize students with government and other data sites and teach them the search skills they need to navigate available resources.

In addition to teaching students where and how to locate and access data, they also teach explicitly the tools and techniques for analyzing and visualizing that data. Though the applications vary according to discipline, a few of the tools are used widely among most, such as Excel, Google Sheets, and Powerpoint. Other exploration and analysis tools include SPSS, GSS Explorer, ArcGIS, Strata, Terrset, R, Strata, and Jamovi. Faculty report a preference for open source software and publicly available tools, to make this work as accessible and as low cost as possible for students. Another big factor in the tools and materials of choice is the ease of use for

faculty; most interviewees expressed interest in using newer or more advanced tools but admitted that constraints on their time prevented them from seeking out or learning new tools for teaching with data. This was a very common response, one that is discussed again in the support section of this report.

Pedagogical Significance: Scholarship and Real-World Application

When asked about the pedagogical significance of their use of data in teaching, participants expressed a desire to prepare their students for their future career path, whether that was a path leading to graduate school, or a career path outside of academia. There is a strong feeling from participants that whichever path students are on, it is important to their professional success that they have basic data literacy skills, that they understand how data is used to answer questions in the social sciences. One sociology professor stated that their goal for student learning is for students to understand that "sociology is a social science, based on empirical evidence that is scientifically collected." In other cases, faculty are concerned with their students' understanding that data literacy will benefit them generally in their future careers: "we want them to recognize that the ability to organize, manipulate, and understand data is a skill, a marketable skill that they can use."

There is also a strong desire to see students graduate from college with an understanding of data and how to be "critical consumers of data." A few instructors do exercises with their classes where they show their students data that is untrue, manipulated, or misleading, and guide them through data analysis to enable them to identify potential problems and recognize faulty interpretations of data, whether good or bad. This type of instruction includes recognizing qualities of good data as well as good sources of data, and metadata.

This pedagogical intentionality extends to the tools faculty ask students to use. Some instructors require that their students use software or programs that are discipline specific and will be used in graduate programs, such as ArcGIS or SPSS. Other instructors, particularly those teaching introductory level courses or to classes where there is a mix of majors and non-majors, use more general tools. One faculty member uses Google maps in assignments for his introductory geography class instead of more advanced GIS software because the students are generally already familiar with it. While in some cases the intention is to spend less time teaching software that is more advanced than necessary for the assignments, in other cases the intention is to ensure that students are able to use software they are likely to be asked to use in their future in any context, such as Excel, which is widely used in both academic and non-academic settings. In some cases, students are asked to access data through publicly available sites which gives the students practice not only in using those sites, but also exposure to the various types of information that is publicly available.

Student and Faculty Challenges: Teaching, Learning, and Re-Learning

Students don't necessarily show up to the first day of class equipped with the skills and knowledge to work with data. Even if they have learned applicable skills and knowledge before, they don't necessarily transfer those skills into their current courses. Many participants find themselves teaching skills or tools with which they assumed their students would already be familiar. One participant stated that each year she discovers more and more computer skills that

she must explicitly teach in her courses, in addition to the course material. Some participants struggled more than others with the balancing of computer and data skill instruction and course material. Whether due to differences in teaching experience, personality, their own comfort using relevant software, or the nature of their course objectives, some participants more easily integrated computer skills into their courses with refreshers early in the semester or instruction at the point of need throughout their courses. Other instructors found it challenging to fit in necessary instruction related to specific skills and software in addition to the discipline specific course material they planned to teach. For example, a prevalent theme in interviews was that of students' lack of comfort in Excel. While many participants mentioned this as one of the most important programs for students to master, some participants were surprised by the amount of time they spend in their classes teaching students to use the software, at the expense of regular class material. One faculty member said that each semester, a number of her students need to be taught the most basic aspects of using Excel, such as how to highlight a row or average a column of data.

There are other concerns related to curriculum and data literacy. Participants are often surprised that their students seem to be lacking basic data literacy skills that the participants had assumed or hoped were being taught earlier or elsewhere in the curriculum. For example, one professor noted that students who have taken statistics don't know how to apply those skills to their archaeology class. In some cases, this lack of transfer is happening even in a course sequence within a discipline. One participant said that in an advanced Geography course, he structures his instruction to include review of the software and other data skills that were covered in the introductory course that students were required to take before the advanced course.

These challenges are compounded by students' fear of math in general and data specifically. There is a feeling from some of the participants that students are surprised to find quantitative data being emphasized in the social sciences. Participants also noticed a fixed mindset in their students, a belief that they are "bad at math" and that data or anything involving numbers is something they are unable to learn. Students display a lot of anxiety when looking at data sets or graphs, which contributed to their difficulty in learning the lessons the instructor was trying to teach. One sociology professor noted that she sees a lot of her students become uncomfortable and withdrawn when they are asked to interpret charts and graphs. While she considers these types of activities to be relatively basic and stepping stones to more complex data related tasks, the student's discomfort becomes an impediment to their learning because, as she says, "if they're not comfortable, they're just not going to really engage."

Faculty also faced challenges in teaching with data. Because students often had fewer data literacy skills than faculty were assuming they would, faculty found they were spending more time teaching the tools and lower level skills involved with data literacy, which ate into time they would have otherwise spent on more advanced skills.

Participants sometimes felt limited by their own skill levels. While they have particular areas of expertise or comfort with some sources of data, they don't always have the time to become familiar with all of the available data sources or emerging technologies. This can hamper their ability to teach content involving data, help students seek out data for particular projects or show students new technology related to data that might be useful for them in their projects or future

courses. Again, many faculty members are autodidacts, often reluctant ones. Without any widespread institutional offerings, they feel forced to learn new skills and tools on their own and on their own time.

Support for and from Colleagues: Sharing and Collaborating

In general, social sciences faculty are community-minded when it comes to making their teaching materials openly available for others, however, it varies widely from the personal to the departmental level. Some departments employ some method of sharing, from informal swapping of syllabi or lessons to a departmental shared drive to a more formal repository arrangement. In one unit, a maternity leave arrangement has allowed a faculty member a course release in exchange for the creation of new materials for teaching with data in a newly-approved mid-level methods course. Teaching circles, facilitated by the university's teaching and learning center, allow for another venue in which faculty who use data as part of their courses have shared things like SPSS guides, syllabi, and lesson ideas.

Most often mentioned among content shared by and with others are datasets, websites, and other OA material. Time is precious among these instructors, and finding just the right dataset or sites offering the kind of data they find useful takes an enormous amount of time and effort. Unless instructors believe themselves to be the sole practitioner of a method or activity, most acknowledge the scarcity of quality resources and are willing lenders and active borrowers of information.

External Support for Teaching: Wants, Needs, and Realities

Most interviewees talked at least some about the difficulty they face meeting course learning outcomes while simultaneously teaching students the data-related skills required to do the work to achieve those course learning outcomes. As discussed in the challenges section, faculty find themselves using course time earmarked for disciplinary content teaching Excel functions or data literacy content they assume students enter class with, particularly in courses with data- or statistics-related prerequisites. Most understand, in the case of first-year or survey courses, that students may not have any experience in data analysis or tools for that purpose and embed this content into their syllabus, but almost every interviewee expressed a wish for a basic skills course that could prepare students to enter classes with a minimum level of capability, ready to tackle social science research without losing time struggling with the common tools of the discipline. As one early-career professor put it, "I just can only spend so much time on creating a bunch of tutorials."

Most expressed interest in more widely available, easily accessible training for students on tools like Excel or SPSS, that would eliminate the need for them to include such instruction each semester. In interviews, a pattern was noticeable, that one would expect: faculty learn over time to adapt to and explain the inability of their students to retain and transfer knowledge gained in previous courses. They adjust for what they recognize as a gap and create methods to integrate necessary skills into their courses. They also acknowledge the frustration they feel in having to engage in what seems repetitive and unplanned work.

At present, GVSU does not offer a credit-bearing information literacy or library skills course, although it does seem that such a course could be a reasonable place to include basic data searching, Excel, and basic data literacy skills. This is something that library faculty discuss regularly, and as yet, have not made a coordinated effort to design, whether in a stand-alone format or integrated into any one discipline. A primary concern with moving in that direction is, again, the recognition that, removed from a specific context, students are unlikely to internalize these skills and concepts and transfer them to other situations or uses. More useful perhaps, would be a coordinated scaffolding of content within each major or discipline, where students encounter basic skills early on, and re-engage with data collection, interpretation, analysis, visualization, and other competencies again and again throughout their advancing coursework. Librarians can still play a crucial role in this process, but in a more integrated manner.

Time and lack of skill or familiarity with new tools and technologies aren't the only issues that leave faculty wishing for external support. All recognize the need to stay informed and prepared, but admit that maintaining even a minimum level of knowledge about new sources of data and new methods for working and teaching with data requires time, discipline, and curiosity. Adding this new learning to an already jam-packed workload can become to some faculty members a Sisyphean undertaking, especially for those less familiar with newer technologies. One interviewee prefaced their own wish for more university-supplied educational support this way: "People have trained at different times and under different circumstances. So as the data needs are...evolving and changing, that sometimes outpaces the training or the technical skills of the folks who are doing the instruction."

Finding data sets for instructional purposes is, as stated before, a time-consuming endeavor, and many rely on the same sets they've used for longer than they'd like because of the labor involved with searching for new ones. Several interviewees would find a centralized repository of all the available data the most useful resource of any they could imagine. Short of that, many would be very happy to just have a way to stay more informed about what data is available and where.

Several faculty members discussed in depth about why they do or don't have their students engage in the collection of data from human subjects. Because of the severe limitations of having their students use family, friends, and student colleagues as subjects, most choose to create mock data sets, or find existing data sets. Those who do have their students collect data from outside of class acknowledge the practice's limits but want their students to experience the collection and coding process. Some participants expressed a desire for a pool of participants for data collection activities, so that students can learn the data collection process without sacrificing the integrity of their student research.

On a more comprehensive level, the need for better coordination and integration of fundamental data-related skills across the general education curriculum was mentioned by a number of the interviewees. Several talked about the need for curricular change, for the scaffolding of data skills within and across courses so that students might experience a more intentional learning process and thus, build, practice, and ideally, transfer what they learn in one course to another.

Recommendations

Based on findings, there are a number of ways in which liaison librarians, library administration, and the greater university community can enact support for faculty data use and instructional goals and help build students' data literacy skills.

Liaison Librarians: Support from the Front Lines

Librarians can help faculty and students locate publicly available data sources for use in courses and faculty research, though there continues to be some debate among library faculty regarding the efficacy, usefulness, and scalability of specific skills or tool workshops facilitated by librarians. Are library faculty the right people to be teaching Excel to students in large-enrollment prerequisite courses? Without a specific disciplinary connection, do students retain and transfer the skills taught in such workshops or courses? On the other hand, the library could be part of a coordinated, collaborative effort, providing the campus with a solid, sustainable source of support for such knowledge and skill building, creating a bridge between multiple disciplines and making that elusive idea of skill transfer more accessible and explicit.

- Liaisons can compile publicly available datasets and tutorials for navigation, and curate for general and specific uses in subject guides or other interactive collections.
- The Libraries' Scholarly Publishing team, in collaboration with liaison librarians, are ready to consult with social science faculty to identify, locate, and curate extant data repositories appropriate for course and research use. Liaison librarians can promote such collections to faculty and students via general mass communications, class visits, departmental meetings, and one-on-one consultations and casual conversations.
- Future and continued conversations with the University's General Education Committee's library representative can ensure that as student learning goals and assessments are adjusted according to need, the Library remains an integral part of supporting data instruction and integration.
- Librarians can train student peer consultants in the Libraries' Knowledge Market specific skills and/or tools so that they can act as "front line" assistance for students in need.

Library: Coordinated Support for Data Literacy

- Develop a coordinated marketing or communication strategy for the many different areas of data literacy support. This strategic approach would detail specific types of support, identify appropriate staff and faculty contacts, and offer suggestions for collaboration and instructional & research support. A multi-modal communication program would include Library Website, email, social media, and face-to-face communication.
- Partner with Library's Knowledge Market and Digital Skills peer consultants to collaborate with social science faculty on data-specific and large-scale assignments, to take some of the burden off of faculty and practice or reinforce repetitive but necessary technical skills, such as Excel, SPSS, or online statistical calculators.

University: Strategy and Integration for a Stronger Curriculum

- Incorporate data literacy concepts and specific skills that were mentioned as participants as especially important and/or daunting to students, such as Excel, into a first-year experience course.
- Provide data software/tools learning opportunities for faculty via the University's Faculty Teaching and Learning Center. Offer incentives for training, such as course release, course overload pay, or credentialing.

Appendix A

Hello,

We are GVSU librarians involved in a national library research project coordinated by [Ithaka S + R](#), about the use of quantitative data in instruction within the social sciences. For the purposes of this project, we are defining data as including research that analyzes physical specimens, landscapes, text, or images using computational methods, including GIS, and in teaching that encourages students to work more dynamically with data.

If you use quantitative data in your instruction, we would very much like to interview you. The semi-structured interview would last for no more than an hour. We understand the constraints on your time through this unusual academic year, and we will try to make this as easy and unobtrusive as possible and work around your schedule.

The aims of this project are several:

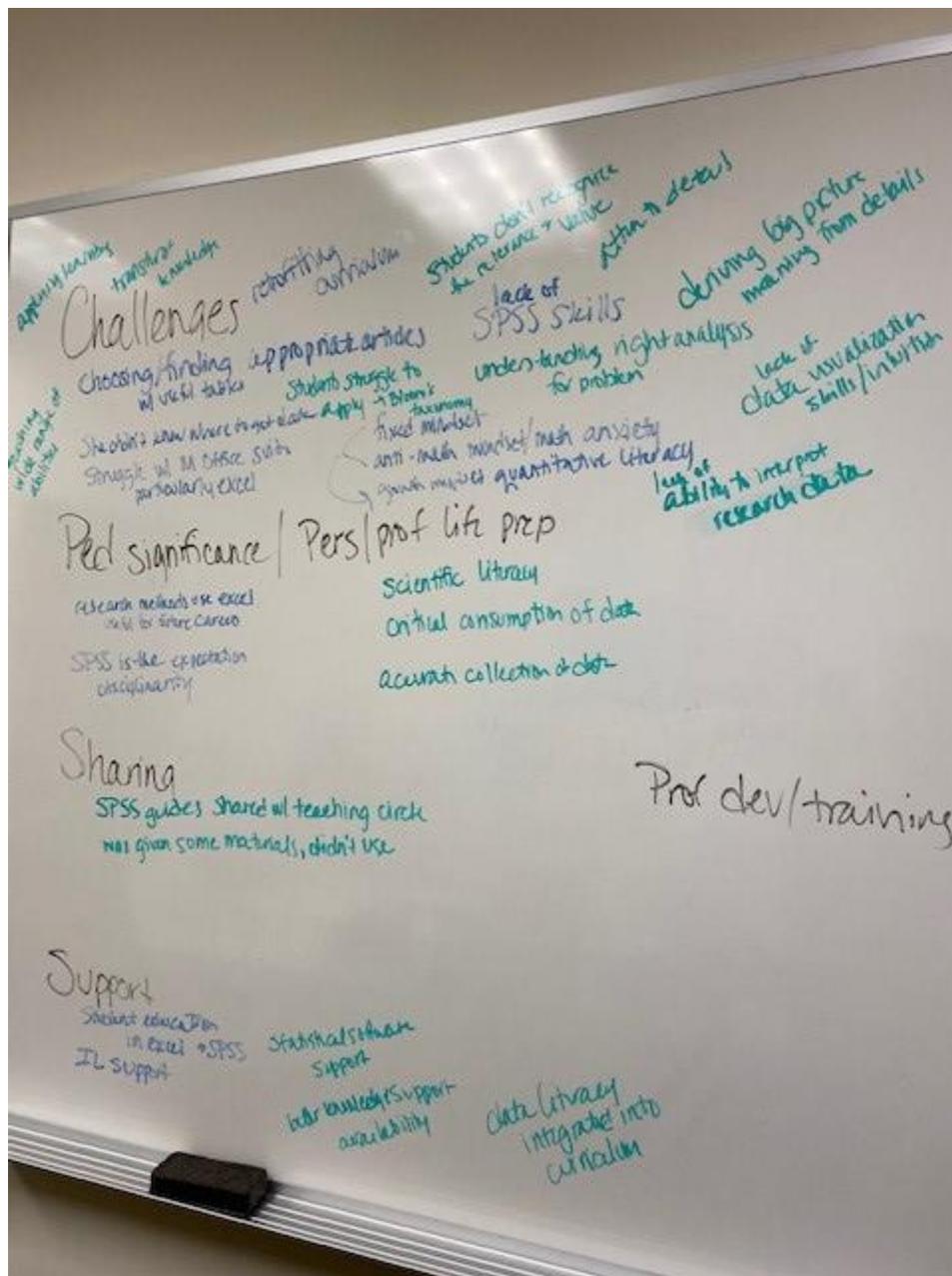
- To deeply explore the ways in which undergraduate instructors in the social sciences teach with quantitative and/or computational data
- To better understand instructors' support needs
- To develop actionable recommendations for faculty, department heads, administrators, and data service providers on campus
- To build or strengthen relationships within Grand Valley and across peer institutions

If you are willing to participate in an interview, your name and any information collected from you or about you is for the sole purpose of this research study and will be kept confidential to the fullest extent allowed by law. The survey has been approved by the IRB (20-310-H).

Thank you for considering! Let us know if you have any questions or need further information.

Gayle Schaub and Samantha Minnis

Appendix B



Example of response categorization and analysis process