

Harvard Data Science Review • Issue 6.4, Fall 2024

Where Data Science and the Disciplines Meet: Innovations in Linking Doctoral Students With Masters-Level Data Science Education

**Doreet Preiss¹ Jessica Sperling¹ Ryan M. Huang^{2,1} Kyle Bradbury³
Thomas Nechyba⁴ Robert Calderbank³ Gregory Herschlag⁵
Jana Schaich Borg¹**

¹**Social Science Research Institute, Duke University, Durham, North Carolina, United States of America,**

²**Conservation Ecology Research Unit, Department of Zoology and Entomology, University of Pretoria, Hatfield, South Africa,**

³**Department of Electrical and Computer Engineering, Duke University, Durham, North Carolina, United States of America,**

⁴**Department of Economics, Duke University, Durham, North Carolina, United States of America,**

⁵**Department of Mathematics, Duke University, Durham, North Carolina, United States of America**

The MIT Press

Published on: Aug 21, 2024

DOI: <https://doi.org/10.1162/99608f92.f81142cc>

License: [Creative Commons Attribution 4.0 International License \(CC-BY 4.0\)](#)

ABSTRACT

Although the need for data science methodological training is widely recognized across many disciplines, data science training is often absent from PhD programs. At the same time, master's-level data science educational programs have seen incredible growth and investment. In 2018, Duke University initiated a National Science Foundation (NSF)-funded program to determine whether master's-level data science programs that universities have already invested in could be leveraged to reduce data science education barriers doctoral students face. Doctoral fellows from diverse fields worked with teams of master's students from Duke's Master in Interdisciplinary Data Science program on applied capstone projects focused on the doctoral fellows' own disciplines and dissertation research. Fellows also gained access to the master's program's courses and professional development resources. We examined the implementation, experience, and effect of this integration into Master in Data Science program infrastructure using qualitative data collection with doctoral fellows, master's students, and fellows' doctoral advisors. Master's students participating in doctoral-led capstones benefited from their doctoral fellows' mentorship, project management, and content knowledge. Participating doctoral students showed increased learning of data science techniques and professional skills development. While some fellows' research was advanced through the capstones, data also showed mismatches between selected master's program goals and doctoral students' needs. Overall, this pilot indicated potential promise in harnessing existing Master in Data Science programs to bolster doctoral students' data science learning and professional readiness while also identifying areas for improving future such efforts.

Keywords: data science, doctoral education, master's education, interdisciplinary, capstone, collaborative research

1. Introduction

The field of data science education has seen incredible growth over the past decade, with new programs and initiatives being developed and honed across major universities ([Dominici et al., 2022](#); [Li et al., 2023](#); [Raj et al., 2019](#); [Zaugg et al., 2021](#)). Despite the proliferation of data science programs and interest in their curricula, most PhD students in more traditional disciplines do not receive adequate training in the applied data wrangling necessary to handle the volume and variety of data collected by both academic and nonacademic entities, nor the soft skills required to succeed in nonacademic settings ([Denecke et al., 2017](#); [Sarkar et al., 2016](#)). This is especially problematic as an increasing number of doctoral graduates enter nonacademic fields where data science acumen is expected, partially due to decreasing availability of academic research positions, with fewer than 17% of STEM PhD students receiving tenure track jobs. For computer science and related fields, that percentage was only 12% in 2018 ([Heflinger & Doykos, 2016](#); [Larson et al., 2014](#); [Petrie et al., 2017](#); [Sinche et al., 2017](#); [Zweben & Bizot, 2019](#)).

Duke University developed the National Science Foundation (NSF)-funded Master in Interdisciplinary Data Science–Innovations in Graduate Education (MIDS-IGE) fellowship program in 2018 to test whether the investment many universities have already made in data science master’s programs could be leveraged to help address the important barriers doctoral students face to data science education and professional readiness. The MIDS-IGE program did this by evaluating the potential benefits of integrating PhD students into select parts of the new (at the time) Duke Master in Interdisciplinary Data Science (MIDS) program. Specifically, doctoral students from varied academic programs joined the MIDS capstone experience as Innovations in Graduate Education (IGE) fellows who collaborated on capstone projects focused on their own dissertation research. IGE fellows informed the scoping and management of their capstone projects, contributed to analyses, led project meetings, and provided deep disciplinary-specific knowledge to guide the capstone work. In turn, they were intended to benefit from the MIDS students’ data science knowledge, their own direct involvement in these projects, opportunities to enroll in core data science courses that were otherwise only available to MIDS students, and opportunities to participate in MIDS professional development programming. These activities were designed to support the overall programmatic goals for PhD students to improve their technical and quantitative skills, gain experience and training in teamwork and project management, gain confidence in their ability to learn and engage in data science work more generally, have the opportunity to engage in experiences relevant to nonacademic careers, and advance their academic research. In addition, MIDS students were intended to benefit from the domain and content expertise of their IGE doctoral student partners. Based on qualitative data collection with IGE fellows, master’s students, and fellows’ doctoral advisors, this study examines the experiences of IGE fellows’ themselves, as well as those of MIDS students who worked on IGE capstones. It also provides a comparative lens through additional data collection with MIDS students who worked on traditional (non-IGE partnered) capstones. Drawing on an empirical study of the implementation and results for this innovative program’s participants, this article can inform ongoing efforts to integrate data science into doctoral training across disciplines.

2. Background: Data Science Graduate Education and the MIDS-IGE Fellowship Program

2.1. Data Science Graduate Education: The Gap Between Skills and Demand

Amid the remarkable expansion of programs and initiatives in data science education across numerous major universities ([Dominici et al., 2022](#); [Zaugg et al., 2021](#)), gaps persist between the skillset possessed by data science students and the expertise sought by industry. Alan Garber, the provost of Harvard University at the time, captures part of the challenge when he states, “the pervasive use of the term ‘data science’ in academic settings reflects both the appeal of the intellectual activities it encompasses and the capaciousness—or vagueness—of its meaning” ([2019](#)). This complexity poses a significant challenge in dedicated data science education programs: how does one cover this multiplicity of areas with sufficient depth while still producing a well-rounded data scientist who can provide impactful value to a future employer ([Hardoon, 2021](#))? As noted

by [Kolaczyk et al. \(2020\)](#), existing postsecondary programs in core data science fields often “leave students, upon exiting academia, needing a nontrivial ramping-up period before they can truly have an impact with their first employers.”

While dedicated data science programs face challenges in integrating interdisciplinary focus, doctoral education in non-data science disciplines faces a distinct yet related challenge: most PhD students in non-data science programs whose work could benefit from data science knowledge do not have sufficient data science training or opportunities to receive such training. For instance, doctoral students are not receiving adequate training in the applied data wrangling necessary to handle the volume and variety of data collected by both academic and nonacademic entities.

Recently, the NSF research community emphasized the need to harness data science in various subjects including chemistry, biology, astronomy, and physics ([NSF, 2020](#)). This need is also echoed in non-STEM fields; with improved quantity and quality of data, researchers in the social sciences and humanities are increasingly encouraged to rely on data science techniques as well ([Adhikari et al., 2021](#); [Pawlicka, 2017](#)). There is also a growing recognition that data science education both informs and is informed by varied fields of study like biomedical sciences, humanities, social sciences, and other traditional fields ([Lazowska, 2018](#); [Pawlicka, 2017](#); [Wing & Banks, 2019](#)). Therefore, these gaps in skills training not only inhibit individual students in their careers, whether in industry or government, but also limit academic fields overall.

Doctoral students interested in building their data science skills and professional competencies often feel their advisors will not support such endeavors and that research assistantships funding their graduate education discourage taking advantage of relevant professional development opportunities ([Woolston, 2015](#)). Moreover, doctoral advisors, many trained for research careers themselves before the advent of data science, often lack training in how to cultivate data or professional skills in their trainees or may feel that the demands of academic research do not leave enough time to manage interdisciplinary projects or train graduate students in nontraditional academic skills ([Gamse et al., 2013](#)). These challenges can result in students who are inadequately prepared for the big data management and statistical computing often necessary for both industry jobs and core academic research and who may lack soft skills that can be particularly critical in interdisciplinary work both within and outside of academic settings ([Denecke et al., 2017](#); [Sarkar et al., 2016](#)).

2.2. Innovations in Data Science Education

Specific developments in graduate data science education that focus on practical applications can help to address these gaps for both dedicated data science degree programs (including undergraduate and master’s level) and doctoral students in other fields whose work would benefit from data science. Students who engage with practical problems through data science projects gain valuable experience in addressing complex real-world challenges ([Kolaczyk et al., 2020](#)), collaborating with domain experts outside of their own disciplines

([Zaugg et al., 2021](#)), and translating their findings into actionable solutions ([Dominici et al., 2022](#)). By emphasizing real-world contexts, such initiatives bridge the gap between theory and application.

Academic programs have developed specific curricular and programmatic components to take advantage of the benefits of this kind of practical experiential learning. This includes using capstone projects, in which students work on synthesizing and applying their knowledge and skills to a real-world problem. One way of implementing capstones is to pair teams of students with outside partners or stakeholders who need help with a data analysis task. These partnerships enable data science students to gain experience working with messy, real-life data while also having the experience of working with clients; in turn, clients benefit from the contributions of the data science students. Numerous data science master's programs employ this approach via capstone or capstone-like projects ([Allen, 2021](#)). While typically assigned at the end of a period of study (e.g., Columbia University's or New York University's [NYU] Data Science master's programs), capstones may be integrated into academic programs at varied points ([NYU, 2024](#); [Zaugg et al., 2021](#)). For example, Boston University (BU's) Master of Science in Statistical Practice (MSSP) has experimented with integrating capstone projects at the start of the program and aligning foundational course learnings with project milestones throughout students' academic journey ([Kolaczyk et al., 2020](#)). In tandem with applied data science training, many programs have added focus on essential real-world soft skills required for professional success, such as communicating to nontechnical audiences (e.g., University of Hong Kong's Master of Data Science program), giving oral presentations (e.g., BU's MSSP), teamwork (e.g., the University of North Carolina [UNC]-Chapel Hill Master of Applied Data Science), and reflecting on nuanced problems from various perspectives (e.g., Harvard's Data Science Initiative) ([Dominici et al., 2022](#); [Kolaczyk et al., 2020](#); [Yu & Li, 2021](#); [UNC-Chapel Hill, 2024](#)).

2.3. Leveraging Data Science Master Programs to Bridge Gaps in Doctoral Education

Despite the proliferation of data science master's programs and corresponding capstones and professional development programming, doctoral students are generally limited in opportunities to benefit from such infrastructure ([Denecke et al., 2017](#); [Woolston, 2015](#)). In addition, although industrial settings are dominated by team-based projects, traditional PhD programs require students to display individual leadership ([Denecke et al., 2017](#); [Woolston, 2015](#)). Therefore, the skills needed to work in a collaborative environment may happen by circumstance rather than by requirement in a traditional PhD program. Adding to the challenge, doctoral programs in data science are themselves limited, with very few offered in the United States ([Casola, 2020](#)). Students looking for specialized training in this space often rely on more traditional PhD programs like statistics or computer science, which as will be discussed in the Results section, can often feel intimidating or inaccessible to students not coming from these backgrounds. The Duke MIDS-IGE program aimed to leverage the resources and infrastructure of the already-developed Master in Data Science program to help address the educational barriers facing PhD students, without demanding much additional infrastructural support from home departments or disciplines. By enabling access to curricular components of the MIDS program, PhD

students may be able to gain many of the applied data science and related soft skills needed to succeed in modern academic and nonacademic settings.

2.4. The MIDS Program

The MIDS program at Duke University was launched in fall 2018 to build on and advance graduate data science education. This program specifically highlights interdisciplinarity within data science, reflecting the fact that data is collected in almost every domain and embracing the idea that data science requires the integration and synthesis of expertise and skillsets from many fields. Building off [Cao's \(2017\)](#) definition, MIDS conceives of data science as what happens when the intersection of statistics, informatics, computing, effective communication, social sciences, and strong team management is applied to the analysis of complex data sets that require deep domain knowledge and critical thinking to interpret.

MIDS students are expected to demonstrate experience in, and passion for, data before applying and have diverse backgrounds in both quantitative and nonquantitative fields. Cohorts have ranged between 31 to 48 students with nearly half of them men and half of them women. To date, the program has had 11% underrepresented students by race and ethnicity (African American/Black and Hispanic/Latino students). Notably, incoming MIDS students have spanned a wide age range (from 21 to 45 years of age), as many data science students already possess a graduate degree or more than 6 years of professional work experience. Housed within Duke's Social Science Research Institute and primarily funded by program tuition, the MIDS program continues to train data scientists and welcomed its seventh cohort of master's students in summer 2024.

A Duke MIDS degree requires 2 years to complete, and includes an online statistics, math, and programming review before Year 1 and a summer internship between Years 1 and 2. In total, MIDS students take eight core courses that cover the diverse technical and nontechnical competencies needed to solve data problems in real contexts (see Figure 1 for the components of the MIDS program). Core courses are designed with teaching techniques that encourage a 'growth mindset' and use problem-based active learning team activities. Students can also choose up to six elective courses from other departments to expand their knowledge in areas of particular interest. Students must also attend workshops to aid in their professional development, including applying data science to business problems, presenting to nontechnical team members, and networking in the data science environment after graduation. MIDS offers activities to prepare students for job interviews in diverse career paths (such as working on elevator pitches and practicing brain teaser problems they may have to solve during interviews) and conducts exercises to develop emotional competencies that are instrumental in achieving professional success, such as curiosity, persistence, emotional intelligence, moral character, and a commitment to continuous learning ([Brown & Ferrill, 2009](#); [O'Boyle et al., 2011](#); [Romanelli et al., 2006](#)).

<u>Analytical Competency</u>	<u>Professional Development</u>	<u>Practical Experience</u>
Online summer review of statistics, linear algebra, and programming	Taught via workshops in Data Science Dialogues seminar, Capstone assignments, and personal coaching	Required internship between Year 1 and Year 2
Core Courses: <ul style="list-style-type: none"> • Data to Decision • Modeling and Representation of Data • Principles of Machine Learning • Data Scraping, Mining, and Analysis • Data Management • Data Visualization, Logic, and Storytelling • Data Science Ethics • Data Science Dialogues Seminar 	Example Topics Covered: <ul style="list-style-type: none"> • Creating measurable success metrics • Stakeholder analysis and management • Creating and adhering to timelines • Project management frameworks • Personal development plans • Navigating professional disagreements • Skills for working in diverse teams • Interview skills 	Projects with real, messy data sets in core courses 1-year long Capstone Project
4-6 electives* Workshops covering additional topics, such as: <ul style="list-style-type: none"> • Asking “the right” questions • Version control using Git 	Example capacities cultivated: <ul style="list-style-type: none"> • Embracing failure as a means for learning • Motivation for life-long learning • Adherence to personal moral values 	

Figure 1. Components of the Duke Master in Interdisciplinary Data Science (MIDS) program.

Like several other data science programs, MIDS includes a capstone project in its second year. In contrast to programs with shorter capstones (e.g., Columbia University, NYU), MIDS requires a year-long capstone during which MIDS students engage in projects that typically involve nonacademic partners from industry, government, nonprofits, or the community. MIDS leverages existing relationships with community and industry partners to foster these projects and collaborates with other Duke entities (e.g., Fuqua School of Business) to establish relationships with companies interested in projects across different programs at Duke. Students are matched to projects based on their stated interests, backgrounds, skillsets, and constraints given by the outside partners. Capstone projects are completed in teams of two to four students who are selected to ensure that the team members have complementary, but diverse, strengths and to ensure the team has all the skills needed for their assigned project. These projects ensure MIDS students have mentored experiences in real-world contexts by the time they graduate and provide an opportunity to learn how to work with real, messy data sets that are burdened by practical data analysis problems, such as largeness of scale, extreme imbalance, mixed sources, missing data, and privacy concerns (Cao, 2017). The capstones also cultivate students’ ability to define their analytical goals and choose appropriate methodologies. These aspects of data science are reportedly the hardest to learn, and often absent from traditional academic programs in math, statistics, and computer science (Hicks & Irizarry, 2018; Hopper, 2015).

The capstone curriculum incorporates various structured elements to ensure the successful application and further development of professional and technical aspects of MIDS students' training from Year 1. Capstone projects are completed over two semesters, with dedicated classroom time and scheduled assignments focused on establishing team expectations, creating work plans with external partners as clients, providing weekly updates, conducting 360-degree evaluations of team performance, and reflecting on management issues. These elements were chosen for their reported contribution to previously described service-based and team-based projects (Aldag & Kuzuhara, 2015; Schwering, 2015). Teams were also provided technical resources and support, including git repositories, secure data storage, cloud computing, and project management tools (e.g., Slack, Trello, Notion, etc.). During the 2019–2020 and 2020–2021 project years, students wrote weekly reports on what they had accomplished in the previous week and what they were planning to do in the coming week. These reports pointed to elements in the repositories (including code, generated data, figures, and other results) and explained methods of scientific inquiry and experimental design. The reports were directly reviewed by the capstone directors and allowed them to flag potential issues. During the 2021–2022 project year, weekly or biweekly meetings replaced formal weekly reports. IGE capstone teams met with their doctoral advisors and non-IGE teams met with their faculty mentors. (See Figure 2 for detailed information on the capstone program design.)

<u>Capstone Project Support / Resources</u>	<u>Assignments & Output</u>	<u>Project & Team Evaluation</u>
<p>Dedicated classroom time and scheduled assignments over two semesters, focused on: establishing team expectations, creating work plans with external partners, providing weekly updates, conducting 360-degree evaluations of team performance, and reflecting on management issues</p> <p>Technical resources and support, including git repositories, secure data storage, cloud computing, and project management tools (e.g., Slack, Trello, Notion)</p> <p>Check-ins with project managers/faculty advisors: <i>Traditional capstones</i> - weekly or biweekly reports/check-ins with project managers (Duke faculty or post-doctoral students) <i>IGE capstones</i> – weekly check-ins with project managers (2019–2020); IGE fellows' doctoral advisors (2020–2022)</p> <p>Meetings with clients: <i>Traditional capstones</i> – meetings with external clients, cadence determined by clients <i>IGE capstones</i> – weekly meetings with external clients (2019–2020) or IGE fellows' doctoral advisors (2020–2022)</p> <p>Capstone team meetings: Regular team meetings to share domain expertise and collaborate on weekly workflows, analyses, papers, repositories, presentations, and other work products</p>	<p>Weekly assignments, including weekly reports on what students accomplished in previous week and next steps for coming week</p> <p>Mid-semester presentation</p> <p>Repositories of code for replication, corresponding documentation, and relevant datasets used in the work</p> <p>Other work products</p> <p>Final presentation</p> <p>Final white paper</p> <p>Potential research papers, dissertation chapters (for IGE capstones)</p>	<p>Review of weekly reports by capstone course directors with feedback provided (2019–2020, 2020–2021); weekly reports were not required in 2021–2022 and were replaced by regular meetings with capstone faculty mentors</p> <p>Feedback on mid-semester presentation, with opportunities to iterate</p> <p>End-of-term review of: *Paper *Repository *Presentation</p> <p>Team Peer review (2020–2022)</p> <p>End-of-term survey (including open-ended questions and questions with Likert scale ratings) completed by capstone faculty mentors, clients (traditional capstones), and IGE fellows' doctoral advisors (IGE capstones)</p>

Figure 2. Vertically integrated data science capstone projects: program design.

Each MIDS capstone team received weekly or biweekly oversight from project managers (2019–2021) or faculty mentors (traditional capstones)/doctoral advisors (IGE capstones) (2021–2022), aligning with each project's research interests and expertise. These project managers, mentors, and advisors supported each team by providing feedback on analyses, advising on how to adapt to challenges, and ensuring engaged and effective communication with participating entities, as research has shown its critical role in generating capstones with valuable outcomes ([Campbell & Lambright, 2011](#)). Project managers were typically postdoctoral students from diverse departments; project mentors and doctoral advisors are Duke faculty.

Through a collaborative effort, each capstone team must achieve a specific outcome for their participating outside partner and present a final presentation with an accompanying white paper about the outcome's implications. The final deliverables undergo evaluation by a panel of MIDS core faculty and outside partners on multiple dimensions, including students' effective communication to diverse audiences, computational strategy, and creativity.

2.5. The MIDS-IGE Program

The idea for PhD student involvement in capstones occurred concurrently with the MIDS program development. Thus, the IGE program was launched in the same year as the MIDS program. The capstone experience proposed within the IGE program is 'vertically integrated,' incorporating team members with different levels of expertise (master's-level, PhD-level, postdoctoral-level, and faculty-level). The concept of vertically integrated capstones was first piloted in undergraduate construction and engineering projects in other institutions ([Mills & Beliveau, 1999](#)) but has been expanded through many programs at Duke. The MIDS-IGE program began with the objective that participating in data science capstone projects would offer PhD students from diverse disciplines the chance to engage in interdisciplinary team-based projects with real-world relevance beyond academia. These projects would enable these doctoral students to gain invaluable experience and acquire the essential skills needed for successful collaboration and problem-solving in professional settings. For participation in the program, a request for proposals was sent out to various departments across Duke University. Applicants were asked to propose a capstone project that was related to their dissertation research, could benefit from data science techniques, and that could potentially involve a nonacademic partner (outside partners were not required after the first IGE fellow cohort). IGE fellows were then selected from the applications received. Selected IGE fellows were provided \$3,000 in research/travel funds. In addition to their IGE project work, fellows participated in the year-long MIDS capstone course and were expected to complete all coursework and activities related to their capstones during the academic year; they also received access to audit or take MIDS core classes for credit, which is not otherwise available to doctoral students at Duke. Additionally, they were invited (but not required) to participate in any other professional development activities provided by the MIDS program, such as job fairs, interview preparation sessions, and seminars with nonacademic parties who use data science.

Depending on the cohort, IGE fellows served as either team members or managers on their capstone projects, which in both cases entailed informing the scoping and management of their projects, providing deep disciplinary-specific knowledge and direction, and practicing project management skills, such as identifying and distributing tasks and tracking group progress. Fellows were also intended to benefit from the MIDS students' data science knowledge through their own direct involvement in these projects, as well as through the aforementioned opportunity to enroll in MIDS core data science courses. MIDS students were intended to benefit from the domain and content expertise of the IGE fellows with whom they were partnered. (See Figure 3 for an abbreviated IGE Program logic model.)

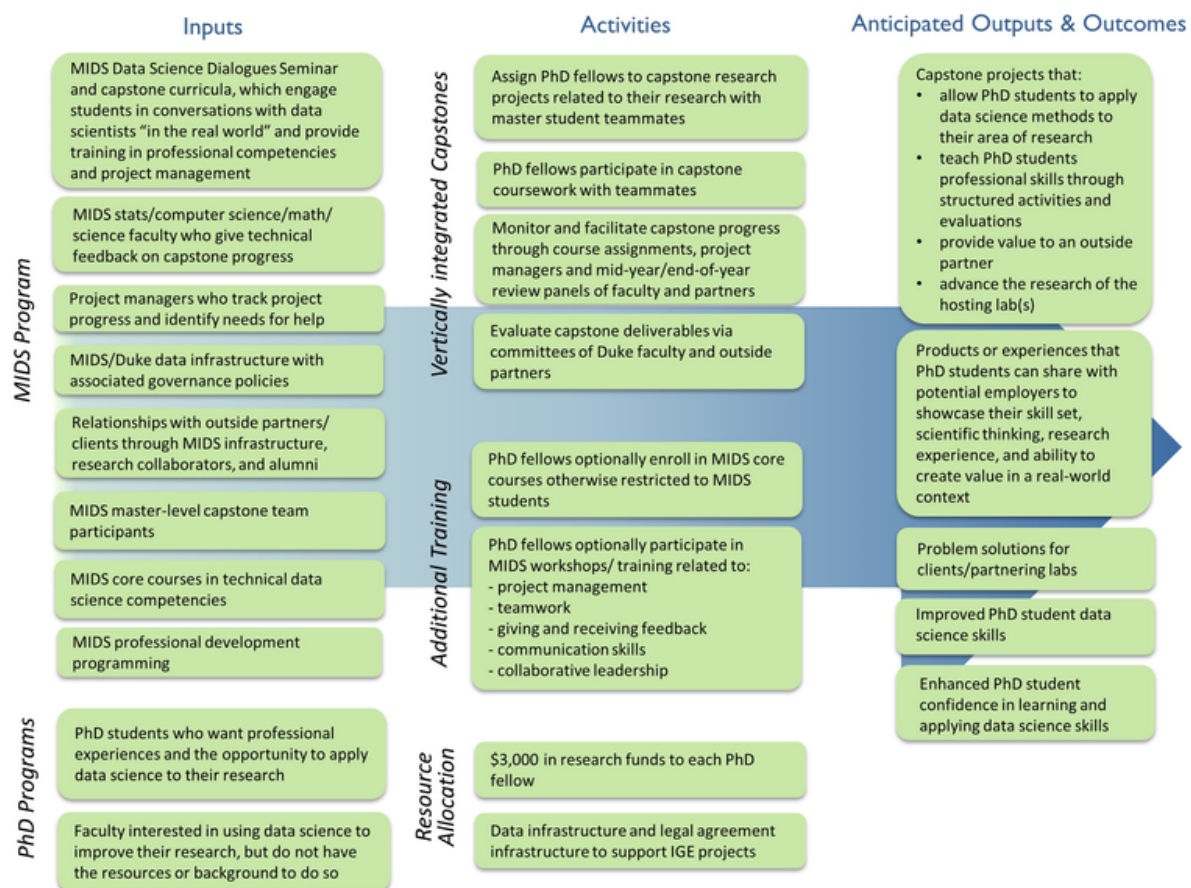


Figure 3. Abbreviated logic model from the IGE Program.

During the program's first 3 years (2019–2020 through 2021–2022), also the years of this study's focus, eight doctoral students participated as IGE fellows. Fellows came from a diverse set of departments, including Sociology, Structural Biology & Biophysics, and Marine Science & Conservation. Two MIDS students participated on each IGE capstone team (one team had a third student), leading to 17 out of 113 total MIDS students partnering on IGE capstones during this time. Since the years of this study's focus, the IGE program has continued to leverage the MIDS capstone infrastructure beyond the initial grant-funded pilot, though the

program is currently recruiting fewer fellows for future cohorts as the MIDS program considers other options for IGE program funding support.

3. Methodology

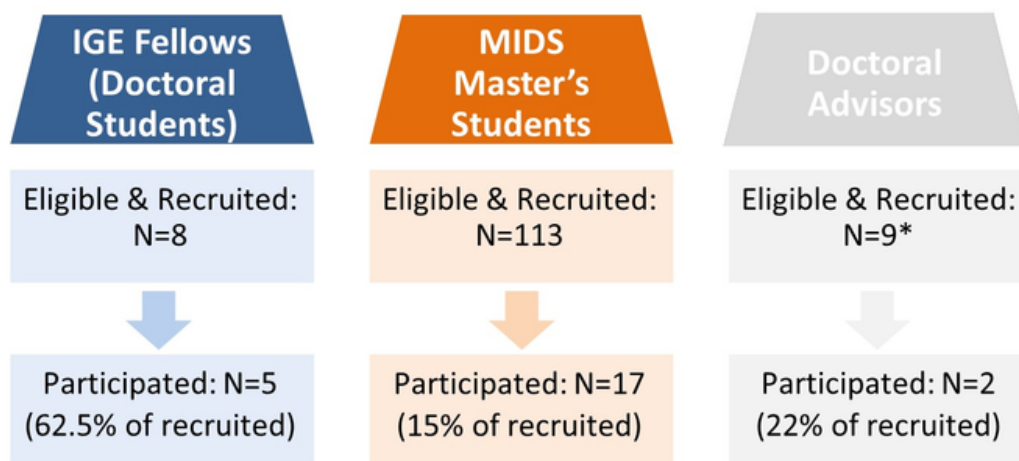
Starting in 2018, the MIDS program developed a relationship with Duke's Social Science Research Institute's Applied Research, Evaluation, and Engagement (SSRI-AREE) team to study the effect of this novel program and consider adjustments that may enhance program effectiveness. This developed into a qualitative evaluation study addressing the IGE program, at that point completing its third year.¹ Although the evaluation team collaborated with MIDS staff to understand program context and development of qualitative instruments, MIDS staff did not take part in data collection or analysis phases of this study; findings were presented to MIDS staff in aggregate after completion of data collection and analyses.

This study addressed two main research questions, with additional subquestions, as outlined below:

1. What is the experience of IGE fellows?
 1. What are key facilitators or barriers for engagement and positive experience on an IGE capstone?
 2. What factors inhibited IGE fellows' experiences?
 3. What effects or outcomes result from engaging with MIDS students on these projects? What factors facilitate or inhibit these effects?
 4. How do IGE fellows' doctoral advisors' view IGE engagement and potential IGE student benefit? What factors do they view as facilitating or inhibiting engagement and value?
2. How do MIDS students experience working on an IGE capstone?
 1. What effects or outcomes do MIDS students experience as a result of engagement in IGE capstones?
 2. What factors contribute to or detract from intended capstone effects and MIDS students' experiences?
 3. Which of these experiences are specific to participation in *IGE capstones* versus *non-IGE capstones* (i.e., traditional capstones with an industry partner)?

These questions were addressed using qualitative semistructured interviews and focus groups with three participant populations: IGE fellows, IGE fellows' doctoral advisors, and current and former MIDS students (both those who participated in IGE capstones and those who worked on traditional, industry-partnered capstones). These three groups were included to address our research questions from an array of participant perspectives. For all respondents, MIDS personnel conducted initial outreach via email, and SSRI-AREE team members followed up with potential participants with a direct invitation to participate, with up to three emails sent to recruit participants. Where extenuating circumstances limited study participation via interviews, an opportunity for written feedback using structured questions based on the interview guide was provided. In total, this study included 24 respondents among the three participant populations. Figure 4 presents the number of fellows, doctoral advisors, and MIDS students who were eligible/recruited and who participated in this research.

For IGE fellows, a semistructured interview guide addressed IGE fellows' thoughts on conducting their research with MIDS students, including whether they perceived benefits or challenges from working on a capstone project while participating in the program. In total, five of the eight fellows, representing all three IGE cohorts, participated. Four fellows participated in approximately 60-minute individual interviews; one participant provided responses via the written response option. IGE doctoral advisors were included in this research to provide an additional lens on program implementation and fellows' experiences. A semistructured interview guide addressed perceptions of their IGE fellows' participation in the program, including whether they would recommend participation to future doctoral students, as well as their own experiences advising an IGE fellow. One out of nine doctoral advisors completed an interview, which lasted approximately 60 minutes; one additional advisor provided responses via the written option.² See Appendix A for the full interview and focus group guides.



*One IGE fellow had two advisors.

Figure 4. Eligibility, recruitment, and participation in the evaluation research.

For current and former MIDS students, a focus group structure was chosen given the value of direct dialogue and exchange between MIDS student participants and to build upon the team-based structure of their engagement within the MIDS program and their capstone projects. This process included separate focus groups with MIDS students who worked on IGE-partnered capstones and those who worked on non-IGE projects to understand differences between these two different capstone experiences. All 113 current and former MIDS students were invited to participate in focus groups. MIDS students were later provided with an additional opportunity to participate in individual interviews if they could not participate in a focus group due to scheduling constraints. Ultimately, 17 MIDS students responded to our requests to participate.³ Thirteen students participated across a total of five focus groups; an additional four students participated in individual interviews. Six of these participants worked on IGE capstones; the other 11 worked on traditional industry-partnered capstones.

Participants in both the IGE and non-IGE groups represented a cross-section of all three participating MIDS cohorts (school years 2019–2020 through 2021–2022), and focus groups lasted approximately 90 minutes. The focus groups and interviews probed students' experiences with their capstone, including questions related to working with specific actors on their project (e.g., their MIDS teammates and the capstone directors), factors contributing to their experiences working on the project, project outcomes, and any ongoing benefits from their capstone participation. MIDS students who worked on IGE capstones were also asked specific questions about working with an IGE fellow as well as working on a more research-oriented project, where applicable.

Due to the COVID-19 pandemic, all data collection occurred remotely via Zoom or through written feedback sent as an email attachment. With the respondents' permission, interview data were recorded and transcribed. Data were analyzed using NVivo, a specialized qualitative data management and analysis software. Evaluation researchers developed a coding schema based on the program logic model and on inductive consideration of this study's data, including both thematic and descriptive coding; see Appendix B for the coding schema. Analysis used a thematic content analysis approach; it addressed key constructs and themes emerging from the data, frequency of reference, relationships between key constructs, and difference by respondent characteristics. All processes related to this data collection were reviewed by Duke University's Institutional Review Board (IRB) and approved through an IRB protocol.

4. Results

Results below are discussed for each of the primary respondents: IGE fellows, IGE fellows' doctoral advisors, and MIDS master's students.

4.1. IGE Fellows

Understanding of Program:

- Motivated to learn and apply data science techniques to their research
- Desired to engage data scientists (MIDS students) in their work

Values/Benefits:

- Learned how to scope data science work
- Learned to apply data science skills to their work
- Built increased confidence in their data science capabilities
- Increased familiarity with project management tools/skills
- Gained experience leading a team
- Received technical assistance relevant to their research
- Produced work that may be continued post-IGE fellowship/graduation

Challenges/Concerns:

- MIDS students' still-developing data science skills
- Structural requirements of capstone program
- Combined roles as team members and managers/leaders
- Role as intermediaries between MIDS students and capstone directors
- IGE projects often supplemental to main research aims

Figure 5. Main findings of the IGE fellows.

Figure 5 above provides a summary of the main findings that emerged from interviews with IGE fellows. The IGE program drew doctoral students based on their desire to learn and apply data science techniques to their research in ways that would otherwise be inaccessible to them, as well as the opportunity to directly engage data scientists (MIDS students) in their work. The IGE fellows who participated in this study cited exposure to data science modeling and data management techniques that could be applied to their research data as the main motivators for participation. From figuring out ways to automatically classify data to analyzing large amounts of ‘messy’ data, IGE fellows joined the program with a desire and an expectation to learn to use data science to optimize their work. As one fellow stated: “I feel like I knew a lot about my research specifically, (...) I had a lot of data, but I didn’t really know about the tools or the best ways to analyze it.” Another fellow was using a big and messy data set, necessitating a system for standardization:

so, in this case, our end product was adding a data science tool to an existing macro, or bigger project that I was a part of (...) but we were having problems with like some data were weekly, some were monthly, some were daily and so we wanted to create more standardized data and we thought that we could use machine learning as a tool for that data.

Although fellows recognized the need for better techniques to analyze their data, their doctoral programs did not offer resources to learn the relevant data science techniques. Therefore, they were enthusiastic about the opportunity to gain relevant data science skills or help through the IGE fellowship.

In addition to the lack of availability through their own departments, the MIDS program also seemed more accessible to these doctoral students, who do not already have data science backgrounds, particularly compared to other data science learning opportunities. One student described considering doctoral-level computer science classes but being too intimidated because she did not already have a baseline knowledge in the field. To her, master's-level work felt less intimidating. Importantly, MIDS courses are only available to MIDS students, with the exception of IGE fellows. Thus, access to MIDS courses was one of her main motivators for applying to be a fellow. Several additional fellows reported auditing one or two MIDS courses on subjects relevant to their work to further their understanding of applying data science techniques to their research.⁴

In addition to gaining access to master's level-data science courses, fellows cited a desire to learn from and work with MIDS students who already had a year of data science coursework experience. Across interviews, we found that fellows entered the program with an expectation that the MIDS students on their projects would have a good foundation and a level of expertise in data science skills. As one student described:

[they] would have a good sense of what methods are out there for analyzing big data sets of text and images (...) and they would be able to help me kind of figure out okay, what can we use because I don't know? And it's the case that [fellows' discipline] tends to be a bit behind the times in terms of (...) methods of doing like NLP (...) to help you analyze these big data sets.

Another student mentioned the perceived opportunity to work with “data science wizards.” Overall, fellows came to the program with a sense that they would bring the questions, content knowledge, and possibly data, and MIDS students would be able to operationalize the questions and then analyze the data using data science methods learned as part of their master's studies. Leaning on the expertise of their MIDS students would, in turn, help move their work forward more efficiently than if they relied on data management and analysis techniques typically used, but often not explicitly taught, in their fields of study. Although the view of MIDS students as resources was among the main motivators for program participation, all interviewed fellows also mentioned a desire to mentor students. While some fellows already had quite extensive mentoring or teaching assistant experience and enjoyed that type of role, others wanted the opportunity to build those skills.

IGE fellows perceived various benefits from their program participation, including learning new data science skills, building confidence in their data science abilities, and producing products they could use in later work. Fellows cited using natural language processing and machine learning techniques; they also mentioned specific tools and packages they employed, including GitLab and Python. Beyond the specific methods cited, though, fellows noted an increased comfort with integrating data science in their work. One student explained that she

was afraid to use machine learning prior to becoming an IGE fellow but became comfortable enough to employ it and market her skills going forward:

before I started I was totally afraid to even look at machine learning (...) and I'm now using it in like several other projects and I know that's gonna be like, it's super cliché, but in [academic discipline] they are looking toward more people who are well versed in data science skills which is why I started out in the first place and if you put like machine learning on anything then people are like ooh, fancy, science and then it's very hireable.

Another fellow mentioned that she routinely references the specific code her capstone team developed and manipulates it in her current work, while another leveraged the IGE fellowship experience to finish a dissertation chapter. One student has since published an academic paper (with MIDS students as coauthors) and has also produced a website based on the IGE work. Other fellows noted that they hope to revisit their IGE work at another point.

Fellows also noted that working as a team helped build their confidence in data science. One fellow mentioned that her team often Googled to figure out how to employ particular methods; others mentioned watching YouTube videos with their teams. Realizing that even data science students had to search for resources and tutorials to identify and employ these techniques gave IGE fellows the confidence that they could do the same. One student described this realization as “validation” that empowered her to break perceived barriers to data science. Becoming comfortable with learning about and exploring advanced data science techniques was a particularly valuable outcome for these doctoral students, making data science feel more approachable and thus opening the door to future data science learning.

In addition, several claimed that their MIDS teammates helped them realize or “scope out” what was possible with data science techniques, since some of them came in with the impression that virtually anything is possible with machine learning. As one student put it, her participation in the program helped bring her goals and expectations “down to earth.” Additionally, one fellow’s MIDS students had deep knowledge of a particular program and helped her troubleshoot her code.

Though IGE fellows reported clear benefits, they also indicated numerous challenges. For instance, they often expected MIDS students to have greater data science skills than was the case, and some felt that they missed an opportunity to learn from their MIDS teammates. In addition to expectations regarding expertise, many IGE fellows believed MIDS students would have more time and be more committed to the project than they were; they also expected MIDS students would be more self-sufficient, taking initiative and anticipating next steps for the project. In reality, fellows often needed to provide detailed timelines and next steps for their teammates. As one respondent explained, “I did not expect to have to delegate weekly tasks and then check up on student’s progress. I was expecting a lot more autonomy from the students.”

Relatedly, IGE fellows also indicated misalignment and challenges regarding their own role. Several indicated their own time commitment exceeded their expectations coming into the program, noting their role often extended into time-consuming aspects that they were not anticipating and for which they were not sufficiently prepared, such as managing team dynamics. Beyond these management challenges, IGE fellows often expressed a sense of broader confusion regarding their role. Some fellows began the program with the impression that they would be team leaders; they expected to present their MIDS students with their research questions, background knowledge, and data before having them take the reins. Others expected to be more hands-on with organization and planning. However, some fellows found themselves as equal partners in the work more often than they anticipated, with this dual role leading to varied challenges. As explained by one fellow: “[the experience] ended up being a lot of switching back and forth between setting expectations and then being not a force of authority, but sort of in that role, but then also having to be on the team.” Other fellows described a similar tension between these team-member and team-leader roles, noting the complexity of inferred hierarchy that comes with being a PhD student working with master’s students. Further complicating the role confusion, some fellows were younger or less experienced than their MIDS teammates, who often had significant work experience before joining MIDS.

IGE fellows also shared instances where capstone program structures and requirements did not necessarily reflect their project’s needs. They reported that regular mandatory timesheets requiring a set number of project hours per week, meant to ensure students’ project engagement, did not necessarily match the ebb and flow of research. Moreover, they were concerned that MIDS students were being graded based on the tasks they completed as noted on the timesheets, but seemingly ‘easy’ tasks (e.g., classifying sound or image data) could take many more hours than the capstone directors may have thought they should take. In addition to the timesheets, fellows cited other requirements, such as developing a story/planning board in order to project plan for the year, as not useful to their work. Finally, fellows expressed similar concerns about the capstone course sessions more generally. While most respondents valued learning how to communicate technical aspects to more general (i.e., nonacademic) audiences, they thought the course content was more relevant to capstone projects with industry partners than to their own work. Due to these mismatches between capstone requirements and IGE project needs, several fellows reported serving as a buffer or intermediary between their MIDS students and the capstone directors regarding capstone requirements that they did not feel worked well with their projects.

When reflecting upon their IGE experiences, IGE fellows generally expressed that the fellowship was valuable; however, they did not feel it was necessary. Further, although some felt the fellowship advanced their dissertation work, others reported that their engagement ultimately did not advance their most direct research or dissertation work as expected, with some feeling the fellowship might have slowed their academic progress. Some noted feeling side-tracked from their main research due to a mismatch between the program requirements and the exigencies of their own research. In some cases, this was due to MIDS aims being outside of disciplinary context, which is often important in doctoral research: “Throughout the year, I had difficulty

keeping the project and its objective firmly rooted in [discipline-specific] theory, which is where it needed to be for it to be part of my research.” In addition, the structural elements of the capstone were not aligned with her needed outputs: “the course was aimed towards writing and producing a data science project, which I found to be in conflict (...) with the framework of an academic research paper.” This respondent continued to explain the differing needs of parties and the role of these varied needs in creating this challenge:

[MIDS students] do not particularly need to know or care about these theories beyond this project and working on an academic paper would not have been helpful for them. But this disconnect in desired outcome [sic] often created difficulty in communicating and agreeing upon what the research question was, what data we needed to answer it, which variables were important, and how to measure them. To be honest, I eventually sort of gave up and the project morphed into something that was doable, but pretty far from the question I was interested in.

While this student noted she was proud of the final paper and the work that went into it, she still felt that capstone requirements detracted from her academic research and that she would not be able to use the work in her dissertation or future projects. Similarly, more than one respondent described how they triaged their work, paring down their plans to focus on what was necessary for MIDS students to complete the capstone project, write the white paper, and earn their master’s degree, even if that meant the IGE fellows would not achieve their own research aims.

4.2. IGE Fellows’ Doctoral Advisors

Understanding of Program:

- Limited understanding of program
- Expected the capstone project to enable fellows to acquire additional data and pursue an academic publication

Values/Benefits:

- Fellows conducted data science work that would not have otherwise been possible

Challenges/Concerns

- Fellows’ opportunity costs to program participation
- Misalignment between doctoral needs and MIDS requirements
- Time needed to manage projects and mentor students

Figure 6. Main findings of the doctoral advisors.

Figure 6 above provides a summary of the main findings of IGE fellows’ doctoral advisors. The doctoral advisor respondents indicated limited knowledge of the IGE program. When asked, these doctoral advisors had a general understanding that the program matched talented students in data science with doctoral students in

order to pursue a research question using data science techniques, ideally with an industry partner. In addition to this basic premise, one respondent had the expectation that participation would enable his fellow to acquire additional data and pursue an academic paper: “there was at least the story that this could yield a data set and a path toward an academic paper that he could not have pursued otherwise.” In contrast, the other respondent, completely unfamiliar with the program until his advisee decided to apply, expressed that he did not have any program expectations.

Both doctoral advisor respondents reported that their doctoral students conducted data science work that would not have been possible without the program. However, both also indicated overarching concerns about opportunity costs. This was largely due to concern about misalignment between academic research needs and MIDS goals, focused on two factors: time needed to manage projects and mentor students and a MIDS product not addressing a research output. This echoes IGE fellows’ reporting on lesser dissertation research alignment than expected. As one explained, doctoral students in his program are extremely research-focused, needing to publish several articles to be competitive on the job market; therefore, any time spent on work not aimed at publication became an opportunity cost. Echoing interviews with IGE fellows, this advisor noted that his fellow’s project needed to be adjusted for the MIDS students to complete their capstone requirement, lessening the direct academic contribution for this IGE fellow. Despite recognizing potential benefits, this advisor noted he was “wary” of these MIDS capstone projects, including the “program feature” that required fellows “to have to kind of go above and beyond to make sure that you’re looking out for them [MIDS students].” While he felt like the final product would not have been produced without IGE participation, he claimed that he would not want another advisee to participate unless the perfect project fell into a student’s lap *or* the IGE program shifted to be less industry- and more research-focused. In contrast, the other advisor interviewed said he would support future students participating in the program; however, he did not believe his student could replicate the data science work completed, despite the program’s goal of enabling doctoral students to learn data science techniques.

4.3. MIDS Master's Students

Understanding of Program:

- Little initial awareness of IGE program

Values/Benefits:

- MIDS students on IGE capstones received valued guidance and support from IGE fellows; such guidance/support was often lacking for students on non-IGE capstones
- Increased likelihood of producing an academic publication

Challenges/Concerns:

- None noted

Figure 7. Main findings of the MIDS master's students.

Figure 7 above summarizes the main findings that emerged from data collection with MIDS master's students. For MIDS master's students, there was often not a clear intention to engage with an IGE fellow or related expectations at the outset. When asked if students considered the differences between IGE and non-IGE capstones at the time of ranking project choices, MIDS students expressed that they were not fully aware of the differences. While some said they realized that some of the choices they ranked were with a PhD student based on the project descriptions, others said they did not realize this until they were assigned to and began their project.

Yet, even without clear expectations, MIDS master's students provided clear evidence of benefit from IGE engagement. Most notably, MIDS students who worked on IGE capstones described a level of support that was lacking in discussions with students on non-IGE capstones. MIDS students working on IGE capstones often referred to the support, guidance, and resources their IGE fellows provided throughout the capstone year. First, they noted that IGE fellows provided extensive data with which to work, a contrast with non-IGE projects for which lack of data and time spent trying to access data were among the chief frustrations noted by MIDS students across multiple non-IGE capstones. In addition, IGE fellows had a great knowledge base and helped 'get them up to speed' with the domain-specific information necessary to work on their projects. From providing relevant background information to creating tutorials for their teammates, IGE fellows served as subject matter experts for their teams; this was generally lacking for those with industry partners. In addition to content expertise, MIDS students cited regular communication (meetings, Slack, etc.) with and guidance from IGE fellows. In contrast to this experience, MIDS students on non-IGE capstones more frequently reported

contacts with their industry partners as notably less communicative and often not exhibiting the same degree of interest in the capstone projects as did IGE fellows.

MIDS students on IGE capstones described a level of personal support that students on non-IGE capstones did not seem to experience. The fellows also helped define goals, both at the macro and micro level, and provided helpful feedback along the way. As one student stated, her team's IGE fellow really helped guide them through the process, defining the overarching goals and steps to reach those goals, in ways that students on non-IGE capstones did not experience:

comparing that with my experience talking to other capstone groups, we were relatively clear in terms of (...) what we were supposed to do in the next few weeks (...) we're pretty clear in terms of how exactly the end result would look like (...) [other teams] spent like a very long time just to figure out what exactly the client wants and what exactly the problem would look like (...) what are some things that they can do to actually help the client achieve their goal, and that was a very long laborious back and forth process (...) comparing to these type of projects we knew exactly what the client want [sic] and how the end product could potentially help them achieve their goal.

For IGE capstones that involved a client—for example, a lab—IGE fellows often served as intermediaries or 'bridges' between the MIDS students and the client. Since the IGE fellows had more regular contact with the client, whereas the MIDS students might meet with them once per semester, fellows would bring questions from MIDS students to their partner. They also helped clarify the clients' wishes to the MIDS students when there was any confusion or a need for clarification. Reflecting on a role reported by IGE fellows in their interviews, MIDS students also saw their fellows as valuable intermediaries between the capstone directors and themselves at times. Students on IGE capstones described how IGE fellows often helped sift through feedback from the capstone directors, figuring out how to navigate requirements that did not necessarily make sense in conjunction with their work, at least in their opinions. While some MIDS students on non-IGE capstones noted that their assigned project manager/mentor⁵ helped them navigate and prioritize sometimes-competing expectations from clients and the capstone directors, overall, most MIDS students on non-IGE capstones did not feel they had this level of guidance and support. MIDS students on non-IGE capstones also more frequently indicated not knowing who to go to if they ran into issues or questions, particularly if their industry contacts were not very involved.

Some MIDS students had additional tangible benefits from being part of an IGE capstone, such as becoming coauthors on an academic paper. Notably, we found no evidence of disadvantages to MIDS students assigned to IGE capstones as opposed to non-IGE capstones.

5. Discussion and Conclusion

Recognizing the need to bridge gaps between traditional academic disciplines and data science programs, the MIDS-IGE program was designed to leverage the curricula and staff of the existing Master in Interdisciplinary

Data Science program at Duke University to teach advanced data science techniques, professional skills, teamwork, and project management to doctoral students who may not receive such training in their home departments. In its first 3 years, the MIDS-IGE program showed the potential to harness existing Master in Data Science programs to bolster data science learning for doctoral students across disciplines. Interviews with IGE fellows indicated that the program increased doctoral students' opportunities to learn about and apply data science techniques to their own research. This was accomplished both through access to MIDS data science courses, which are not otherwise accessible to non-MIDs students, and through work alongside their MIDS student teammates. In addition to specific data science knowledge, the IGE program provided fellows with confidence in their data science capabilities. IGE fellows reported employing data science techniques or code developed during the capstone in their subsequent research and listing data science knowledge on their CVs, which they perceived as a boon to their credentials on the job market. Beyond data science itself, fellows also noted an appreciable benefit in developing soft skills, such as working with and managing a team, scoping projects with tight deadlines, and mentoring students.

While the program demonstrated benefits for IGE fellows, it also provided important advantages for MIDS students who worked on IGE capstones. In addition to gaining a deeper knowledge of domain content by working alongside their IGE fellows, IGE-partnered MIDS students noted that the IGE fellows were very present and invested in their capstone projects; fellows themselves also cited this as they were both invested in their own research and in the success of their MIDS students. In contrast, students working on traditional capstones often perceived a lack of investment on the part of their industry partners. IGE-partnered MIDS students also cited open communication lines with their fellows, including assistance from their fellows when capstone requirements seemed unproductive or unrealistic; industry-partnered MIDS students did not feel that they had a similar ally and often cited unrealistic expectations from their industry partners themselves.

Although the MIDS-IGE program provided clear benefits to participants, there were some drawbacks to participation. As evidenced in interviews with all three respondent groups—IGE fellows, doctoral advisors, and MIDS master's students—there was a relative lack of program knowledge going into the IGE program. This often led to misaligned or unrealistic expectations that colored participants' views of their experiences. It is possible that the stresses and challenges of the COVID-19 pandemic (which took place in the middle years of the MIDS-IGE program) also exacerbated the negative impact of perceived misalignment, but improved communication and transparency with regards to program logistics and expectations could help better match candidates and their advisors to the program and facilitate smoother experiences for all parties involved. The MIDS-IGE program's main mechanisms for attempting to set expectations were written agreements that outlined the goals and processes of the MIDS-IGE fellow program and discussions with both PhD fellows and their advisors prior to awarding the IGE fellowship and while the capstone was in progress. More experimentation will be needed to determine how best to supplement these efforts moving forward.

Results shared in this article also suggest that future iterations of the MIDS-IGE program at Duke, as well as other universities wishing to implement similar programs, should pay careful attention to alignment between doctoral and master's students program needs. By tailoring the IGE capstone experience to better reflect doctoral research requirements, the program could improve data science knowledge and access for PhD students in traditional disciplines, while alleviating concerns that fellows may get side-tracked from their research agenda. Aspects such as the pace and deliverables of each groups' work should be considered, and modifications may be needed to ensure that doctoral students do not feel like they need to triage or forgo their desired end-products (often, but not always, a dissertation chapter) in order for master's students to meet their program requirements. This is particularly important given that data science capstone projects often have to navigate the process of modifying project objectives while the project is underway due to a lack of suitability of the available data to the partner's original objectives ([Allen, 2021](#); [Mowbray, 2015](#)); such an outcome can be a valuable learning experience for data science students that is common to many data science endeavors, but may be incompatible with the performance expectations placed on PhD students. Program organizers should also be cognizant of the time commitment needed to both mentor students and help them complete program requirements. Minimizing deliverables that may be unnecessary for doctoral students' particular work may enable them to focus their allotted fellowship time on data science skills and mentoring/working alongside their master's student teammates. In addition, as mentor alignment to projects proved very beneficial to everyone in the project, the alignment and investment of mentors (e.g., IGE fellows, faculty mentors and/or project managers, where applicable) should be considered in future program development, where possible.

Beyond program participants' experience and benefit, feasibility of program implementation, including a careful consideration of resources required, is critical when considering the launch and sustainability of such a program. The MIDS-IGE program was specifically designed to leverage the existing MIDS program; creating capstone infrastructure and curricula from scratch requires significant and sustained resources, such as dedicated faculty to monitor progress, mechanisms for recruiting outside partners, data and legal processes to manage proprietary data, resources to aid conflict management and facilitate professional communication, and impartial evaluators to monitor and analyze impact ([Kolaczyk et al., 2020](#); [Maleki, 2009](#); [Mowbray, 2015](#); [Reinicke & Janicki, 2011](#)). Incorporating IGE fellows and IGE capstones added minimal additional burden to the responsibilities of MIDS capstone faculty; other institutions with significant data science infrastructure may be able to incorporate PhD students into their programming with a similar manageable level of extra investment. On the other hand, it would likely be difficult and expensive for an institution without a data science capstone infrastructure to integrate a similar program into its offerings.

There is considerable consensus that experiential learning of the kind achieved through capstone projects is a particularly effective way to teach 'data science thinking' and professional readiness ([Allen, 2021](#); [Hicks & Irizarry, 2018](#)), but each educational institution will need to determine how much to invest in capstone projects given their own educational goals and constraints. More research is needed to determine how to achieve the benefits of the MIDS-IGE program in differing settings, but lessons from the inaugural MIDS-IGE cohorts can

provide a model for increasing data science access to academic disciplines that are increasingly focused on large-scale data, yet do not have the departmental resources to provide their own such training and support.

Acknowledgments

We would like to thank the IGE fellows, doctoral advisors, and MIDS students who participated in data collection for this study and provided invaluable insights into the first years of the MIDS-IGE program. We would also like to acknowledge the contributions of Sarwari Das, M.S., who provided research and writing support for this work.

Contributions

Regarding the evaluation study: DP, JS, and JSB led the conceptualization and methodology development for this evaluation. DP led investigation, including conducting the qualitative data collection; conducted data analysis; and led evaluation project administration with support from JS. DP and JS led the original drafting of this manuscript. RMH, GH, KB, JSB, TN, and RC provided review and editing of the manuscript. JSB let the grant that funded the evaluation. JS provided supervision on the evaluation. Regarding the MIDS and associated IGE programs themselves: TN, RC, and JSB were involved in program conceptualization and funding acquisition. RMH, GH, KB, JSB, TN, and RC were directly involved in implementation of the MIDS program.

Disclosure Statement

The MIDS-IGE program was initiated with the support of the National Science Foundation via grant 1806593. The authors have no financial or nonfinancial disclosures to share for this article.

References

- Adhikari, A., DeNero, J., & Jordan, M. I. (2021). Interleaving computational and inferential thinking: Data science for undergraduates at Berkeley. *Harvard Data Science Review*, 3(2).
<https://doi.org/10.1162/99608f92.cb0fa8d2>
- Aldag, R., & Kuzuhara, L. (2015). *Creating high performance teams: Applied strategies and tools for managers and team members*. Routledge.
<https://www.taylorfrancis.com/books/mono/10.4324/9780203109380/creating-high-performance-teams-loren-kuzuhara-ray-aldag>
- Allen, G. I. (2021) Experiential learning in data science: Developing an interdisciplinary, client-sponsored capstone program. In M. Sherriff & L. D. Merkle (Eds.), *SIGCSE '21: Proceedings of the 52nd ACM Technical Symposium on Computer Science Education* (pp. 516–522). ACM. <https://doi.org/10.1145/3408877.3432536>

- Brown, D., & Ferrill, M. J. (2009). The taxonomy of professionalism: Reframing the academic pursuit of professional development. *American Journal of Pharmaceutical Education*, 73(4), Article 68.
<https://doi.org/10.5688%2Faj730468>
- Campbell, D., & Lambright, K. (2011). How valuable are capstone projects for community organizations? Lessons from a program assessment. *Journal of Public Affairs Education*, 17(1), 61–87.
https://orb.binghamton.edu/public_admin_fac/32
- Cao, L. (2017). Data science: Challenges and directions. *Communications of the ACM*, 60(8), 59–68.
<https://doi.org/10.1145/3015456>
- Casola, L. (Ed). (2020). *Roundtable on data science postsecondary education: A compilation of meeting highlights*. The National Academies Press. <https://doi.org/10.17226/25804>
- Denecke, D., Feaster, K., & Stone, K. (2017). *Professional development: Shaping effective programs for STEM graduate students*. Council of Graduate Schools. https://legacy.cgsnet.org/publication-pdf/4062/CGS_ProfDev_STEMGrads16_web.pdf
- Dominici, F., Langdon-Gray, E., & Parkes, D. C. (2022). Spinning up a data science initiative at Harvard. *Harvard Data Science Review*, 4(4). <https://doi.org/10.1162/99608f92.ad105ec8>
- Gamse, C., Espinosa, L., & Roy, R. (2013). *Essential competencies for interdisciplinary graduate training in IGERT*. Abt Associates. <https://files.eric.ed.gov/fulltext/ED553183.pdf>
- Garber, A. M. (2019). Data science: What the educated citizen needs to know. *Harvard Data Science Review*, 1(1). <https://doi.org/10.1162/99608f92.88ba42cb>
- Hardoon, D. R. (2021). Is data science education a jack of all trades? *Harvard Data Science Review*, 3(1). <https://doi.org/10.1162/99608f92.d1aaf1d7>
- Heflinger, C. A., & Doykos, B. (2016). Paving the pathway: Exploring student perceptions of professional development preparation in doctoral education. *Innovative Higher Education*, 41(4), 343–358.
<https://doi.org/10.1007/s10755-016-9356-9>
- Hicks, S. C., & Irizarry, R. A. (2018). A guide to teaching data science. *The American Statistician*, 72(4), 382–391. <https://doi.org/10.1080/00031305.2017.1356747>
- Hopper, T. (2015, May 11). How I became a data scientist despite having been a math major.
<https://tdhopper.com/blog/how-i-became-a-data-scientist/>
- Kolaczyk, E. D., Wright, H., & Yajima, M. (2020). Statistics practicum: Placing “practice” at the center of data science education. *Harvard Data Science Review*, 3(1). <https://doi.org/10.1162/99608f92.2d65fc70>

- Larson, R. C., Ghaffarzadegan, N., & Xue, Y. (2014). Too many PhD graduates or too few academic job openings: The basic reproductive number R_0 in academia. *Systems Research and Behavioral Science*, 31(6), 745–750. <https://doi.org/10.1002/sres.2210>
- Lazowska, E. (2018, March 4). How to encourage data-driven discovery. *The Chronicle of Higher Education*. <https://www.chronicle.com/article/how-to-encourage-data-driven-discovery/>
- Li, D., Milonas, E., & Zhang, Q. (2023). Rankings vs realities: Exploring competency differences in graduate data science programs. In *2023 IEEE Frontiers in Education Conference (FIE)* (pp. 1–4). IEEE. <https://doi.org/10.1109/FIE58773.2023.10343290>
- Maleki, R. A. (2009). Business and industry project-based capstone courses: Selecting projects and assessing learning outcomes. *Industry and Higher Education*, 23(2), 91–102. <https://doi.org/10.5367/000000009788146647>
- Mills, T., & Beliveau, Y. (1999). Vertically integrating a capstone experience: A case study for a new strategy. *Journal of Construction Education*, 4, 278–288.
- Mowbray, J. (2015). The postgraduate capstone experience: Negotiating the pedagogical tensions. *Journal of Learning Design*, 8(2), 43–52. <https://doi.org/10.5204/jld.v8i2.233>
- National Science Foundation. (2020). *Harnessing the Data Revolution (HDR): Data Science Corps (DSC) - Building capacity for HDR* [Program solicitation – NSF 21-523]. https://nsf.gov-resources.nsf.gov/solicitations/pubs/2021/nsf21523/nsf21523.pdf?VersionId=hiTcG4JrTGQyYkdxcm7OWvgx_7hV0Yk
- New York University Center for Data Science. (2024). *Master's in data science*. Retrieved April 19, 2024, from <https://cds.nyu.edu/capstone-project/>
- O'Boyle, E. H. Jr., Humphrey, R. H., Pollack, J. M., Hawver, T. H., & Story, P. A. (2011). The relation between emotional intelligence and job performance: A meta-analysis. *Journal of Organizational Behavior*, 32(5), 788–818. <https://doi.org/10.1002/job.714>
- Pawlicka, U. (2017). Data, collaboration, laboratory: Bringing concepts from science into humanities practice. *English Studies*, 98(5), 526–541. <https://doi.org/10.1080/0013838X.2017.1332022>
- Petrie, K. A., Carnahan, R. H., Brown, A. M., & Gould, K. L. (2017). Providing experiential business and management training for biomedical research trainees. *CBE Life Sciences Education*, 16(3). <https://doi.org/10.1187/cbe.17-05-0074>

Raj, R. K., Parrish, A., Impagliazzo, J., Romanowski, C. J., Ahmed, S. A., Bennett, C. C., Davis, K. C., McGettrick, A., Pereira, T. S. M., & Sundin, L. (2019). Data science education: Global perspectives and convergence. In B. Scharlau & R. McDermott (Eds.), *ITiCSE '19: Proceedings of the 2019 ACM Conference on Innovation and Technology in Computer Science Education* (pp. 265–266). ACM.

<https://doi.org/10.1145/3304221.332553>

Reinicke, B., & Janicki, T. (2011). Real world projects, real world problems: Capstones for external clients. *Information Systems Education Journal*, 9(3), 23–27.

Romanelli, F., Cain, J., & Smith, K. M. (2006). Emotional intelligence as a predictor of academic and/or professional success. *American Journal of Pharmaceutical Education*, 70(3), Article 69.

<https://doi.org/10.5688%2Faj700369>

Sarkar, M., Overton, T., Thompson, C., & Rayner, G. (2016). Graduate employability: Views of recent science graduates and employers. *International Journal of Innovation in Science and Mathematics Education*, 24, 31–48.

Schwering, R. E. (2015). Optimizing learning in project-based capstone courses. *Academy of Educational Leadership Journal*, 19(1), 90–104.

Sinche, M., Layton, R. L., Brandt, P. D., O'Connell, A. B., Hall, J. D., Freeman, A. M., Harrell, J. R., Cook, J. G., & Brennwald, P. J. (2017). An evidence-based evaluation of transferrable skills and job satisfaction for science PhDs. *PLOS ONE*, 12(9), Article e0185023. <https://doi.org/10.1371/journal.pone.0185023>

University of North Carolina at Chapel Hill. (2024). *The online Master of Applied Data Science program: From UNC School of Data Science and Society*. Retrieved April 9, 2024, from <https://online.unc.edu/online-masters-programs/masters-data-science/>

Wing, J. M., & Banks, D. (2019). Highlights of the inaugural Data Science Leadership Summit. *Harvard Data Science Review*, 1(2). <https://doi.org/10.1162/99608f92.e45fcb79>

Woolston, C. (2015). Graduate survey: Uncertain futures. *Nature*, 526(7574), 597–600. <https://doi.org/10.1038/nj7574-597a>

Yu, P. L. H., & Li, W. K. (2021). Project-based learning via competition for data science students. *Harvard Data Science Review*, 3(1). <https://doi.org/10.1162/99608f92.44f54f00>

Zaugg, I., Culligan, P., Witten, R., & Zheng, T. (2021). Collaboratory at Columbia: An aspen grove of data science education. *Harvard Data Science Review*, 3(4). <https://doi.org/10.1162/99608f92.53c4a1b4>

Zweben, S., & Bizot, B. (2019). 2018 Taulbee survey: Undergrad enrollment continues upward; Doctoral degree production declines but doctoral enrollment rises. *Computing Research Association*, 31(5), 3–74. <https://cra.org/crm/2019/05/2018-taulbee-survey/>

Appendix

Appendix A. Interview and Focus Group Guides

IGE Fellow Interview Guide:

Introductory Prompt

Thank you so much for speaking with me today! I'm [*name*], with Duke's Social Science Research Institute. I will be conducting this research interview today. Before we begin, let me tell you a bit about what we're doing today.

Our team at Duke University's Social Science Research Institute, also known as SSRI, has partnered with the MIDS organizers to better understand how the IGE program has worked for participants and how it can be improved. As part of this effort, SSRI is requesting current and former IGE students participate in a brief interview about their experience. Our goal here is to get a deeper understanding of your experiences in the MIDS-IGE program. What you share will contribute to our recommendations for the program, in terms of improving it for future participants.

With your permission, I would like to record the interview. The recording will be transcribed by a third party. The recording will be retained as long as the project is active and will be destroyed when the research is completed. The audio recordings and transcripts will not be shared outside of the research team. All information you provide will be confidential; your individual responses will not be shared outside of SSRI staff. After talking to you, findings will be reported back to MIDS staff, without any information to identify individual participants. Where relevant, participants will be described based on their role—for example, MIDS or IGE student—rather than specific titles. Where direct quotations are included, we will take care to ensure that these do not contain any potentially identifying information that could be easily linked back to you. The data collected will never be shared outside the Duke University research team who are approved as part of this project by the Institutional Review Board; in addition to understanding program experience and gain, the research team may use this information to inform future research related to data science programs.

Your participation in this interview is voluntary. You can also skip any question you would like. If you want to stop taking part after the interview has begun, just let me know. There are no direct benefits to you for joining this study; however, the information you provide can help inform and improve the MIDS-IGE program.

The format of this interview is fairly informal. I do have specific areas to discuss, but there is no need to keep to a strict question-and-answer format. Also, please be open and truthful in your responses; as a reminder, this is all confidential.

I expect that we'll be talking for about 1 hour, though this could vary depending on the amount of feedback you want to provide.

So, are you comfortable participating? [Yes/no]

Do I have permission to record our interview today? [Yes/no]

Do you have any questions for me before we begin?

Thank you so much for your time and cooperation! We truly value your input.

Motivation for engagement and initial expectations

First, I'd like to ask about your reasons for participating in the program.

- When you participated in the IGE program, you worked on [XXX] project.
 - Can you tell me a little about why/how did you decide to pursue this work?
 - Why did you decide to apply to be an IGE fellow?
- What were your expectations when you began the IGE program?
 - What was your understanding of the primary goals of this program when you initially signed on?
 - What were your initial expectations of your role in this program?
 - What were your initial expectations of the MIDS students on your team?

Experience in program

Now I'd like to shift to think about your overall experience with the IGE program.

- Overall, how did the program go for you?
 - *[Probe for specific program components]*. Specifically, thinking about various aspects of the program. Tell me about your experience with ...
 - The project as a whole?
 - Did you work with an outside partner?
 - How did you find that partner? Were you matched by the IGE program, or did you connect with them?
 - How was that experience?
 - Positives?
 - Challenges?

- The MIDS students who were placed on your project?
 - What did you feel was successful about working with this group?
 - Were there aspects that didn't go so well with this group? What were they?
- Your advisor?
 - What role did s/he play? How involved was s/he?
 - How did you feel about this?
- The capstone director?
 - What do you think was helpful about their advising of your project?
 - Was there anything that didn't work so well? What was it?
- Your own role?
 - How did you feel about your role in the group?
 - What worked well?
 - What didn't work so well?
 - [If being a project manager not mentioned already] Did you serve as a project manager?
 - How did that work (or not work) for you?
- The capstone course?
 - Did you participate? Why or why not?
 - [If participated] Were there any aspects of the course that felt particularly beneficial?
 - Were there any aspects that didn't feel very useful/helpful?

Now, I'd like to think about the expectations you mentioned as we started this interview.

- Did your experience match your expectations?
 - How so?
 - How did it not?

Reflections on benefits/challenges of your IGE experience

[For all IGE interviews]

First, I'd like to ask you to only reflect on your year of participation.

- What aspects of this experience did you feel were **most** beneficial during your participation year? What did you gain from this experience?
 - What aspects of the program do you think contributed most to these gains/benefits?
- What aspects of this experience did you feel were **least** beneficial or most challenging during your participation?
 - What aspects of the programming/experience do you think contributed most to these challenges?
 - What do you think could have been done to make this more beneficial to you?

[For current IGE students]

- Are there key areas that the program hasn't addressed for you (i.e., gaps in the programming)?
 - What would you add/change to address this gap?
- If you had the power to **change** parts of the program for future IGE participants, what would you change, if anything? Why?

[For former IGE fellows]

Now, I'd like to hear your thoughts and reflections since completing your IGE year.

- Since finishing your year as an IGE fellow, do you feel like you have experienced ongoing benefits? What are they?
- Have you had any ongoing challenges related to your IGE participation?
- Are there key areas that the program didn't address for you (i.e., gaps in the programming)?
 - What would you add/change to address this gap?
- If you had the power to **change** parts of the program for future IGE participants, what would you change, if anything? Why?

[For all IGE interviews]

Closing

- Is there anything you'd like to share that we haven't already discussed regarding your IGE experience?

Focus Group Guide for MIDS Students Working on IGE Capstones:

Introductory Prompt

Thank you so much for speaking with me today! I'm [*name*], with Duke's Social Science Research Institute. I will be conducting this focus group today. Before we begin, let me tell you a bit about what we're doing today.

Our team at Duke University's Social Science Research Institute, also known as SSRI, has partnered with the MIDS organizers to better understand how the IGE program has worked for participants and how it can be improved. As part of this effort, SSRI is requesting current and former MIDS students who work/worked on a MIDS-IGE capstone participate in a focus group about their experience. Our goal here is to get a deeper understanding of your experiences with the program. What you share will contribute to our recommendations for the program, in terms of improving it for future participants.

With your permission, I would like to record this session. The recording will be transcribed by a third party. The recording will be retained as long as the project is active and will be destroyed when the research is completed. The audio recordings and transcripts will not be shared outside of the research team. All

information you provide will be confidential; your individual responses will not be shared outside of SSRI staff. After talking to you, findings will be reported back to MIDS staff, without any information to identify individual participants. Where relevant, participants will be described based on their role—for example, MIDS or IGE student—rather than specific titles. Where direct quotations are included, we will take care to ensure that these do not contain any potentially identifying information that could be easily linked back to you. The data collected will never be shared outside the Duke University research team who are approved as part of this project by the Institutional Review Board; in addition to understanding program experience and gain, the research team may use this information to inform future research related to data science programs.

Your participation in this focus group is voluntary. You can also skip any question you would like. If you want to stop taking part after the focus group has begun, just let me know. There are no direct benefits to you for joining this study; however, the information you provide can help inform and improve the MIDS-IGE program.

The format of this focus group is fairly informal. I do have specific areas to discuss, but there is no need to keep to a strict question-and-answer format. Also, please be open and truthful in your responses; as a reminder, this is all confidential; we ask everyone here to keep what is said within this group.

I expect that we'll be talking for about 90 minutes, though this could vary depending on the amount of feedback you want to provide.

So, are you comfortable participating? [Yes/no]

Do I have permission to record our focus group today? [Yes/no]

Do you have any questions for me before we begin?

Thank you so much for your time and cooperation! We truly value your input.

Motivation for engagement and initial expectations

First, I'd like to ask about your reasons for participating in the capstone, and particularly an IGE capstone.

- What project did you work on?
 - Did you choose this project or were you assigned to it?
 - If you chose this project, what interested you about it?
 - Did you have specific feelings about being part of an *IGE capstone*?
- What were your expectations when you began the capstone?
 - What was your understanding of the primary goals of this program when you initially signed on?
 - What were your initial expectations of your role working on this project?
 - What were your initial expectations of the IGE fellow (i.e., PhD student) who you worked with?
 - What were your expectations of the other MIDS students working on your project?

Experience in program

Now I'd like to shift to think about your overall experience working on an IGE project.

- Overall, how did the capstone go for you?
 - *[Probe for specific program components]*. Specifically, thinking about various aspects of the capstone. Tell me about your experience with ...
 - Working on the project as a whole?
 - Working with the other MIDS students on your project?
 - What did you feel was successful about working together?
 - Were there aspects that didn't go so well? What were they?
 - The capstone director?
 - What do you think was helpful about their advising of your project?
 - Was there anything that didn't work so well? What was it?
 - The IGE fellow who you worked with?
 - What role did s/he play?
 - What worked well?
 - What didn't work well?
 - Your own role?
 - How did you feel about your role within the group?
 - What worked well?
 - What didn't work so well?
 - The capstone course?
 - What was helpful or valuable?
 - What didn't work well?
- Did your capstone experience match your expectations coming into the program?
 - How so?
 - How did it not?
 - Do you think anything specific to participating in an *IGE* capstone, above and beyond capstones overall, contributed to this?

Overall Reflections, Benefits, and Challenges

Now, I want to step back and ask you some overall questions, particularly around benefits and challenges.

[For all MIDS students]

First, I'd like to ask you to reflect only on your year of capstone participation.

- Are there particular skills you feel like you gained from participating in the capstone experience that were helpful during that year?
 - What are they? How were they helpful to you during that year?
 - Do you think you gained anything specific from participating in an *IGE* capstone, above and beyond capstones overall?
- What aspects of this experience did you feel were **most** beneficial during your participation year? What did you gain from this experience?
 - What aspects of the program do you think contributed most to these gains/benefits?
- What aspects of this experience did you feel were **least** beneficial or were most challenging during your participation?
 - What aspects of the programming/experience do you think contributed most to these challenges?
 - Do you think these challenges were more present for you because you participated in an *IGE capstone*? In other words, do you think challenges would have been different if you were NOT part of an IGE capstone?
 - What do you think could have been done to make this experience more beneficial to you?

[For current MIDS students]

- Are there key areas that the program isn't addressing for you (i.e., gaps in the programming)?
 - What would you add/change to address this gap?
- If you had the power to **change** parts of the program for future IGE participants, what would you change, if anything? Why?
 - Is there anything you would specifically change about the *IGE* capstone, above and beyond capstones overall?

[For former MIDS students]

Now, I'd like to hear your thoughts and reflections *since completing* your capstone; this is thinking in the *years after you* finished your time with MIDS.

- Since finishing your capstone, do you feel like you have experienced ongoing benefits in your next professional or academic steps? What are they?
 - Do you think you gained anything specific to participating in an *IGE* capstone, above and beyond capstones overall?
- Are there key areas that the program didn't address for you (i.e., gaps in the programming) that you think could have prepared you for your next professional or academic steps?
 - Do you think these gaps were more present for you because you participated in an *IGE capstone*? In other words, do you think gaps would have been different if you were NOT part of an IGE capstone?
 - What would you add/change to address this gap?

- If you had the power to **change** parts of the capstone for future MIDS students, what would you change, if anything? Why?
 - Is there anything you would specifically change about the *IGE* capstone, above and beyond capstones overall?

Closing

- Is there anything you'd like to share that we haven't already discussed regarding capstones, including working on an IGE project for your capstone experience?

Focus Group Guide for MIDS Students Working on Traditional (Non-IGE) Capstones:

Introductory Prompt

Thank you so much for speaking with me today! I'm [name], with Duke's Social Science Research Institute. I will be conducting this focus group today. Before we begin, let me tell you a bit about what we're doing today.

Our team at Duke University's Social Science Research Institute, also known as SSRI, has partnered with the MIDS organizers to better understand how capstone projects have worked for participants, and we would like to hear from MIDS students to learn about their experiences. Our goal here is to get a deeper understanding of your capstone experience. What you share will contribute to our recommendations, in terms of improving it for future participants.

With your permission, I would like to record this session. The recording will be transcribed by a third party. The recording will be retained as long as the project is active and will be destroyed when the research is completed. The audio recordings and transcripts will not be shared outside of the research team. All information you provide will be confidential; your individual responses will not be shared outside of SSRI staff. After talking to you, findings will be reported back to MIDS staff, without any information to identify individual participants. Where relevant, participants will be described based on their role—for example, MIDS student—rather than specific titles. Where direct quotations are included, we will take care to ensure that these do not contain any potentially identifying information that could be easily linked back to you. The data collected will never be shared outside the Duke University research team who are approved as part of this project by the Institutional Review Board; in addition to understanding program experience and gain, the research team may use this information to inform future research related to data science programs.

Your participation in this focus group is voluntary. You can also skip any question you would like. If you want to stop taking part after the focus group has begun, just let me know. There are no direct benefits to you for joining this study; however, the information you provide can help inform and improve the capstone experience.

The format of this focus group is fairly informal. I do have specific areas to discuss, but there is no need to keep to a strict question-and-answer format. Also, please be open and truthful in your responses; as a reminder,

this is all confidential; we ask everyone here to keep what is said within this group.

I expect that we'll be talking for about 90 minutes, though this could vary depending on the amount of feedback you want to provide.

So, are you comfortable participating? [Yes/no]

Do I have permission to record our focus group today? [Yes/no]

Do you have any questions for me before we begin?

Thank you so much for your time and cooperation! We truly value your input.

Motivation for engagement and initial expectations

First, I'd like to ask about your reasons for participating in the program.

- What project did you work on?
 - Did you choose this project or were you assigned to it?
 - If you chose this project, what interested you about it?
- What were your expectations for your capstone project?
 - What was your understanding of the primary goals of the capstone project?
 - What were your initial expectations of your role working on this project?
 - What were your expectations of the other MIDS students working on your project?
- Did you apply to work on an IGE capstone?
 - Why or why not?

Experience in program

Now I'd like to shift to think about your overall experience working on the capstone project.

- Overall, how did the capstone go for you?
 - [*Probe for specific program components*]. Specifically, thinking about various aspects of the capstone. Tell me about your experience with ...
 - Working on the project as a whole?
 - Working with the other MIDS students on your project?
 - What did you feel was successful about working together?
 - Were there aspects that didn't go so well? What were they?
 - The capstone director?
 - What do you think was helpful about their advising of your project?
 - Was there anything that didn't work so well? What was it?

- Your own role?
 - How did you feel about your role within the group?
 - What worked well?
 - What didn't work so well?
- The capstone course?
 - What was helpful or valuable?
 - What didn't work well?
- Did your capstone experience match your expectations coming into the program?
 - How so?
 - How did it not?

Overall Reflections, Benefits, and Challenges

Now, I want to step back and ask you some overall questions, particularly around benefits and challenges.

[For all MIDS students]

First, I'd like to ask you to reflect only on your year of capstone participation.

- Are there particular skills you feel like you gained from participating in the capstone experience that were helpful during that year?
 - What are they? How were they helpful to you during that year?
- What aspects of this experience did you feel were **most** beneficial during your participation year? What did you gain from this experience?
 - What aspects of the program do you think contributed most to these gains/benefits?
- What aspects of this experience did you feel were **least** beneficial or most challenging during your participation?
 - What aspects of the programming/experience do you think contributed most to these challenges?
 - What do you think could have been done to make this more beneficial to you?

[For current MIDS students]

- Are there key areas that the program has not addressed for you (i.e., gaps in the programming)?
 - What would you add/change to address this gap?
- If you had the power to **change** parts of the capstone experience for future MIDS students, what would you change, if anything? Why?

[For former MIDS students]

Now, I'd like to hear your thoughts and reflections *since completing* your capstone; this is thinking in the *years after you* finished your time with MIDS.

- Since finishing your capstone, do you feel like you have experienced ongoing benefits in your next professional or academic steps? What are they?
- Are there key areas that the program didn't address for you (i.e., gaps in the programming) that could have prepared you for your next professional or academic steps?
 - What would you add/change to address this gap?
- If you had the power to **change** parts of the capstone for future MIDS students, what would you change, if anything? Why?

Closing

- Is there anything you'd like to share that we haven't already discussed your capstone experience?

Doctoral Advisor Interview Guide:

Introductory Prompt

Thank you so much for speaking with me today! I'm [name], with Duke's Social Science Research Institute. I will be conducting this research interview today. Before we begin, let me tell you a bit about what we're doing today.

Our team at Duke University's Social Science Research Institute, also known as SSRI, has partnered with the MIDS organizers to better understand how the IGE program has worked for participants and how it can be improved. As part of this effort, SSRI is requesting IGE doctoral advisors participate in a brief interview about their and their students' experiences with the program. Our goal here is to get a deeper understanding of IGE students' experiences in the MIDS-IGE program. What you share will contribute to our recommendations for the program, in terms of improving it for future IGE students.

With your permission, I would like to record the interview. The recording will be transcribed by a third party. The recording will be retained as long as the project is active and will be destroyed when the research is completed. The audio recordings and transcripts will not be shared outside of the research team. All information you provide will be confidential; your individual responses will not be shared outside of SSRI staff. After talking to you, findings will be reported back to MIDS staff, without any information to identify individual participants. Where relevant, participants will be described based on their role—for example, MIDS or IGE student—rather than specific titles. Where direct quotations are included, we will take care to ensure that these do not contain any potentially identifying information that could be easily linked back to you. The data collected will never be shared outside the Duke University research team who are approved as part of this project by the Institutional Review Board; in addition to understanding program experience and gain, the research team may use this information to inform future research related to data science programs.

Your participation in this interview is voluntary. You can also skip any question you would like. If you want to stop taking part after the interview has begun, just let me know. There are no direct benefits to you for joining this study; however, the information you provide can help inform and improve the MIDS-IGE program.

The format of this interview is fairly informal. I do have specific areas to discuss, but there is no need to keep to a strict question-and-answer format. Also, please be open and truthful in your responses; as a reminder, this is all confidential.

I expect that we'll be talking for about 1 hour, though this could vary depending on the amount of feedback you want to provide.

So, are you comfortable participating? [Yes/no]

Do I have permission to record our interview today? [Yes/no]

Do you have any questions for me before we begin?

Thank you so much for your time and cooperation! We truly value your input.

Pre-IGE expectations

First, I'd like to ask about your thinking and expectations before your IGE fellow began his/her project.

- You advised [NAME] on [XXX] project.
 - Did you recommend him/her to apply to be a MIDS-IGE fellow?
 - Why did you/did you not recommend that he/she apply?
- What were your expectations when [NAME] began the IGE program?
 - What was your understanding of the primary goals of this program?
 - In what ways did you think it would or would not be valuable for him/her?

[NAME's] Experience with program

Now I'd like to shift to think about your overall impression of [NAME'S] IGE experience. Note that we are also speaking with the IGE fellows themselves; your lens on their experience can just help to give us additional perspective.

- How much did you know about [NAME'S] experience? How much did s/he share with you about his/her experience?
- Overall, how do you think the program worked for this student?
 - What do you think [NAME] gained from participating as an IGE fellow?
 - What aspects of this experience did you feel were **most** beneficial to [NAME]?

- Did you feel like any aspects of the program were particularly useful? (For example, working with an outside partner or working with the MIDS students on the team). Why?
- What aspects of this experience did you feel were **least** beneficial to [NAME]? Why?
- Do you think [NAME] experienced any particular challenges as a result of participation?
 - What were they?
 - Do you think [NAME] experienced challenges with any particular aspects of the program? (For example, working with an outside partner or working with the MIDS students on the team). Why?
- What do you think, if anything, could have been done to make this more beneficial/less challenging to [NAME]?
- Do you feel that [NAME'S] experience matched your expectations of how this IGE experience would go?
 - How so?
 - How did it not?
- Thinking ahead, would you recommend future doctoral students participate in MIDS as IGE fellows?
 - Why / why not?
 - Are there certain characteristics of a doctoral student that would make them a better fit? A less strong fit?

Your own experience with program

Now, I'd like to know more about your experience advising an IGE fellow.

- How involved were you in [NAME'S] IGE experience? [Ask to elaborate]
- Do you think this IGE program benefited you at all? If so, how?
- Were there particular challenges advising an IGE fellow? What were they?
- Do you feel that your own experience with IGE matched your expectations?
 - How so?
 - How did it not?

Closing

- Are there key areas that you feel the program didn't address that they should for IGE fellows (i.e., gaps in the programming)?
 - What would you add/change to address this gap?
- If you had the power to **change** parts of the program for future IGE participants, what would you change, if anything? Why?
- Is there anything you'd like to share that we haven't already discussed regarding the IGE experience?

Appendix B. Qualitative Coding Schema

Name		Description
Actors		
	Capstone Director	References to the capstone director role
	Doctoral Advisor	References to doctoral advisor role
	IGE Fellow	References to IGE fellow role
	MIDS Students	References to MIDS student role
	Outside Partner	References to role of outside partner
Capstone Experience (including IGE Experience)		
	Challenges-Difficulties	Challenges/difficulties with the capstone experience
	Other Thoughts	Additional thoughts on the capstone experience
	Positives-Benefits	Positives/benefits of the capstone experience
	Suggestions for Improvement	Suggestions for improving the capstone experience
Capstone Project Activities		
	Capstone Course	Description of capstone course, including course activities and thoughts on the course
	Data	Data for capstone project, including references to data access, working with project data, etc.
	Other	Other capstone project activities

	Scope of Work	Capstone project scope of work, including challenges
Expectations		Expectations of the capstone project and IGE fellowship (for IGE fellows)
Quotables		Illustrative quotes for potential use in reporting
Reasons and Motivations for Participating		Reasons/Motivations for participating in the IGE fellowship (for IGE fellows) or working on IGE capstones (for MIDS students working on IGE capstones)
Topics		
	Communication	Communication between program actors (e.g., capstone directors and program participants, IGE fellows and MIDS students)
	Knowledge of IGE Program	Knowledge of the IGE Program (fellowship and capstone projects) prior to participation
	Management Style	Reflections on management of capstone directors, IGE fellows, and outside project partners
	Support	Reflections on support provided by capstone directors, IGE fellows, and outside project partners

©2024 Doreet Preiss, Jessica Sperling, Ryan M. Huang, Kyle Bradbury, Thomas Nechyba, Robert Calderbank, Gregory Herschlag, and Jana Schaich Borg. This article is licensed under a [Creative Commons Attribution \(CC BY 4.0\) International license](https://creativecommons.org/licenses/by/4.0/), except where otherwise indicated with respect to particular material included in the article.

Footnotes

1. In addition to this study, researchers with Duke’s Social Science Research Institute have partnered with the directors of the MIDS program to conduct annual surveys of MIDS students’ experiences in the program. While these have informed program adjustments, this current study represents a specific focus on, and deeper exploration into, the IGE capstone. ↵
2. This research was conducted toward the end of the spring semester, a particularly busy time for faculty, which likely impacted advisors’ participation. Although greater participation would have been ideal, the main focus of this work was to understand the capstone experience from the perspectives of IGE fellows and MIDS students, with advisors’ perspectives serving as an additional lens. ↵
3. For 2019–2020, 2020–2021, and 2021–2022, respective participation was: IGE fellows ($n = 1$, 33%; $n = 3$, 100%; $n = 1$, 100%), MIDS master’s students ($n = 5$, 14%; $n = 7$, 15%; $n = 5$, 17%), doctoral advisors ($n = 2$, 50%; $n = 0$, 0%; $n = 0$, 0%). Percentages represent response rate among eligible participants by cohort. ↵
4. IGE fellows would not need credits from MIDS courses for their doctoral degrees; therefore, IGE fellows could freely choose between auditing or taking MIDS courses for credit. All of them chose to audit. ↵
5. During interviews with students on non-IGE capstones, the titles of ‘mentor’ and ‘project manager’ seemed to be used interchangeably, when they were mentioned, with some students seemingly not knowing the official title or intended role of these support persons. ↵

References

- Adhikari, A., DeNero, J., & Jordan, M. I. (2021). Interleaving computational and inferential thinking: Data science for undergraduates at Berkeley. *Harvard Data Science Review*, 3(2).
<https://doi.org/10.1162/99608f92.cb0fa8d2>
↵
- Aldag, R., & Kuzuhara, L. (2015). *Creating high performance teams: Applied strategies and tools for managers and team members*. Routledge.
<https://www.taylorfrancis.com/books/mono/10.4324/9780203109380/creating-high-performance-teams-loren-kuzuhara-ray-aldag>
↵
- Allen, G. I. (2021) Experiential learning in data science: Developing an interdisciplinary, client-sponsored capstone program. In M. Sherriff & L. D. Merkle (Eds.), *SIGCSE '21: Proceedings of the 52nd ACM Technical Symposium on Computer Science Education* (pp. 516–522). ACM.
<https://doi.org/10.1145/3408877.3432536>

[↵](#)

- Brown, D., & Ferrill, M. J. (2009). The taxonomy of professionalism: Reframing the academic pursuit of professional development. *American Journal of Pharmaceutical Education*, 73(4), Article 68. <https://doi.org/10.5688%2Faj730468>

[↵](#)

- Campbell, D., & Lambright, K. (2011). How valuable are capstone projects for community organizations? Lessons from a program assessment. *Journal of Public Affairs Education*, 17(1), 61–87. https://orb.binghamton.edu/public_admin_fac/32

[↵](#)

- Cao, L. (2017). Data science: Challenges and directions. *Communications of the ACM*, 60(8), 59–68. <https://doi.org/10.1145/3015456>

[↵](#)

- Casola, L. (Ed). (2020). *Roundtable on data science postsecondary education: A compilation of meeting highlights*. The National Academies Press. <https://doi.org/10.17226/25804>

[↵](#)

- Denecke, D., Feaster, K., & Stone, K. (2017). *Professional development: Shaping effective programs for STEM graduate students*. Council of Graduate Schools. https://legacy.cgsnet.org/publication-pdf/4062/CGS_ProfDev_STEMGrads16_web.pdf

[↵](#)

- Dominici, F., Langdon-Gray, E., & Parkes, D. C. (2022). Spinning up a data science initiative at Harvard. *Harvard Data Science Review*, 4(4). <https://doi.org/10.1162/99608f92.ad105ec8>

[↵](#)

- Gamse, C., Espinosa, L., & Roy, R. (2013). *Essential competencies for interdisciplinary graduate training in IGERT*. Abt Associates. <https://files.eric.ed.gov/fulltext/ED553183.pdf>

[↵](#)

- Garber, A. M. (2019). Data science: What the educated citizen needs to know. *Harvard Data Science Review*, 1(1). <https://doi.org/10.1162/99608f92.88ba42cb>

[↵](#)

- Hardoon, D. R. (2021). Is data science education a jack of all trades? *Harvard Data Science Review*, 3(1). <https://doi.org/10.1162/99608f92.d1aaf1d7>

[↵](#)

-

↑

- Hicks, S. C., & Irizarry, R. A. (2018). A guide to teaching data science. *The American Statistician*, 72(4), 382–391. <https://doi.org/10.1080/00031305.2017.1356747>

↑

- Hopper, T. (2015, May 11). How I became a data scientist despite having been a math major. <https://tdhopper.com/blog/how-i-became-a-data-scientist/>

↑

- Kolaczyk, E. D., Wright, H., & Yajima, M. (2020). Statistics practicum: Placing “practice” at the center of data science education. *Harvard Data Science Review*, 3(1). <https://doi.org/10.1162/99608f92.2d65fc70>

↑

- Larson, R. C., Ghaffarzadegan, N., & Xue, Y. (2014). Too many PhD graduates or too few academic job openings: The basic reproductive number R_0 in academia. *Systems Research and Behavioral Science*, 31(6), 745–750. <https://doi.org/10.1002/sres.2210>

↑

- Lazowska, E. (2018, March 4). How to encourage data-driven discovery. *The Chronicle of Higher Education*. <https://www.chronicle.com/article/how-to-encourage-data-driven-discovery/>

↑

- Li, D., Milonas, E., & Zhang, Q. (2023). Rankings vs realities: Exploring competency differences in graduate data science programs. In *2023 IEEE Frontiers in Education Conference (FIE)* (pp. 1–4). IEEE. <https://doi.org/10.1109/FIE58773.2023.10343290>

↑

- Maleki, R. A. (2009). Business and industry project-based capstone courses: Selecting projects and assessing learning outcomes. *Industry and Higher Education*, 23(2), 91–102. <https://doi.org/10.5367/000000009788146647>

↑

- Mills, T., & Beliveau, Y. (1999). Vertically integrating a capstone experience: A case study for a new strategy. *Journal of Construction Education*, 4, 278–288.

↑

- Mowbray, J. (2015). The postgraduate capstone experience: Negotiating the pedagogical tensions. *Journal of Learning Design*, 8(2), 43–52. <https://doi.org/10.5204/jld.v8i2.233>

↑

-

↑

- New York University Center for Data Science. (2024). *Master's in data science*. Retrieved April 19, 2024, from <https://cds.nyu.edu/capstone-project/>

↑

- O'Boyle, E. H. Jr., Humphrey, R. H., Pollack, J. M., Hawver, T. H., & Story, P. A. (2011). The relation between emotional intelligence and job performance: A meta-analysis. *Journal of Organizational Behavior*, 32(5), 788–818. <https://doi.org/10.1002/job.714>

↑

- Pawlicka, U. (2017). Data, collaboration, laboratory: Bringing concepts from science into humanities practice. *English Studies*, 98(5), 526–541. <https://doi.org/10.1080/0013838X.2017.1332022>

↑

- Petrie, K. A., Carnahan, R. H., Brown, A. M., & Gould, K. L. (2017). Providing experiential business and management training for biomedical research trainees. *CBE Life Sciences Education*, 16(3). <https://doi.org/10.1187/cbe.17-05-0074>

↑

- Raj, R. K., Parrish, A., Impagliazzo, J., Romanowski, C. J., Ahmed, S. A., Bennett, C. C., Davis, K. C., McGettrick, A., Pereira, T. S. M., & Sundin, L. (2019). Data science education: Global perspectives and convergence. In B. Scharlau & R. McDermott (Eds.), *ITICSE '19: Proceedings of the 2019 ACM Conference on Innovation and Technology in Computer Science Education* (pp. 265–266). ACM. <https://doi.org/10.1145/3304221.332553>

↑

- Reinicke, B., & Janicki, T. (2011). Real world projects, real world problems: Capstones for external clients. *Information Systems Education Journal*, 9(3), 23–27.

↑

- Romanelli, F., Cain, J., & Smith, K. M. (2006). Emotional intelligence as a predictor of academic and/or professional success. *American Journal of Pharmaceutical Education*, 70(3), Article 69. <https://doi.org/10.5688%2Faj700369>

↑

- Sarkar, M., Overton, T., Thompson, C., & Rayner, G. (2016). Graduate employability: Views of recent science graduates and employers. *International Journal of Innovation in Science and Mathematics Education*, 24, 31–48.

↑

- Schwering, R. E. (2015). Optimizing learning in project-based capstone courses. *Academy of Educational Leadership Journal*, 19(1), 90–104.
[↩](#)
- Sinche, M., Layton, R. L., Brandt, P. D., O’Connell, A. B., Hall, J. D., Freeman, A. M., Harrell, J. R., Cook, J. G., & Brennwald, P. J. (2017). An evidence-based evaluation of transferrable skills and job satisfaction for science PhDs. *PLOS ONE*, 12(9), Article e0185023. <https://doi.org/10.1371/journal.pone.0185023>
[↩](#)
- University of North Carolina at Chapel Hill. (2024). *The online Master of Applied Data Science program: From UNC School of Data Science and Society*. Retrieved April 9, 2024, from <https://online.unc.edu/online-masters-programs/masters-data-science/>
[↩](#)
- Wing, J. M., & Banks, D. (2019). Highlights of the inaugural Data Science Leadership Summit. *Harvard Data Science Review*, 1(2). <https://doi.org/10.1162/99608f92.e45fcb79>
[↩](#)
- Woolston, C. (2015). Graduate survey: Uncertain futures. *Nature*, 526(7574), 597–600. <https://doi.org/10.1038/nj7574-597a>
[↩](#)
- Yu, P. L. H., & Li, W. K. (2021). Project-based learning via competition for data science students. *Harvard Data Science Review*, 3(1). <https://doi.org/10.1162/99608f92.44f54f00>
[↩](#)
- Zaugg, I., Culligan, P., Witten, R., & Zheng, T. (2021). Collaboratory at Columbia: An aspen grove of data science education. *Harvard Data Science Review*, 3(4). <https://doi.org/10.1162/99608f92.53c4a1b4>
[↩](#)
- Zweben, S., & Bizot, B. (2019). 2018 Taulbee survey: Undergrad enrollment continues upward; Doctoral degree production declines but doctoral enrollment rises. *Computing Research Association*, 31(5), 3–74. <https://cra.org/crn/2019/05/2018-taulbee-survey/>
[↩](#)