ASEE 2022 ANNUAL CONFERENCE
Excellence Through Diversity
MINNEAPOLIS, MINNESOTA, JUNE 26TH-29TH, 2022 ASEE
Paper ID #36723

Building the Corps: an HBCU+PWI Partnership to Broaden Participation in Data Science (Experience)

Kathleen Luchini Colbry (Assistant Dean, Engineering Graduate Student Services)

Katy Luchini Colbry is the Assistant Dean for Graduate Student Services at the College of Engineering at Michigan State University, where she completed degrees in political theory and computer science. A recipient of the NSF Graduate Research Fellowship, she earned Ph.D. and M.S.E. degrees in computer science and engineering from the University of Michigan. She has published dozens of peer-reviewed works related to her interests in engineering education and graduate student success, and has been recognized as a Master Facilitator by the National Research Mentor Network. She is also co-PI for two NSF-funded projects that enhance graduate student success: one is a professional development program offering training in communication, teamwork, leadership, and mentoring skills for interdisciplinary work; the other is a joint effort with Spelman College (a historically black institution serving women) to expand opportunities and increase diversity in STEM by developing a five-year BS+MS program. Dr. Luchini-Colbry also volunteers as the Director of the Engineering Futures Program of Tau Beta Pi, the Engineering Honor Society, which provides interactive professional development seminars for thousands of engineering students and professionals each year.

Dirk Joel-luchini Colbry (Research Specilest)

Marta Dark

Vincent Melfi (Associate Professor)

Tiffany Renee Oliver (Dr.)

Building the Corps: an HBCU+PWI Partnership to Broaden Participation in Data Science (Experience)

Abstract

This experience paper describes initial efforts to build the data science workforce by developing a partnership between an historically black college or university (HBCU) and a predominantly white institution (PWI). The HBCU is Spelman College, an institution nationally recognized for excellence in supporting minority students; it offers robust undergraduate programs in biology, physics, and math – but currently has no formal data science program. The PWI is Michigan State University (MSU), a large, research-intensive institution with experience developing data science programs for undergraduates, graduate students, and working professionals seeking to upskill. Faculty from both institutions have been collaborating for about two years to understand the unique strengths and needs of their home campuses and how they can combine efforts to broaden participation and diversify the data science workforce.

The National Academies and the National Science Foundation have recognized the need to build the data science workforce to meet growing demand to leverage "big data" across the economy. While there is certainly a need for experts with a deep knowledge of data science principles and computational skills, many emerging opportunities are seeking individuals with expertise in other disciplines (like natural and life sciences) as well as experience working with data. In order to provide this type of discipline+data training, we must address barriers like instructors who lack the appropriate background or training to teach data science; students who lack the appropriate background or prerequisite courses to support lessons in data science; and a lack of space in the major curriculum requirements.

To help address this need, private and public funding has been obtained for a five-year collaboration that will (1) create opportunities for natural science majors from Spelman College to study data science; (2) equip faculty from both institutions to integrate data science into their teaching and research, with a particular focus on using data to address social justice issues; (3) support mentoring, professional skills training, career development, and community-building activities to foster student success. The project encourages both students and faculty to engage in experiential learning; to build skills for leveraging big data in teaching, research, and industry activities; and to experience the cultures, communities, and opportunities of each campus.

A key component of this effort is a new, five-year BS+MS program that will allow students pursuing a bachelor's degree in natural science from Spelman to also earn a master's degree in data science from MSU. Flexible, modular coursework has been designed to help students build foundational skills and fill knowledge gaps as they move between natural science and data science studies. Spelman students will be able to take individual courses in data science, to complete a new data science minor, or to pursue the BS+MS. This project also includes training and support to help faculty integrate data science into their classrooms and labs, and to build skills for inclusive mentoring. Ongoing research collaborations, shared seminars, and faculty/student exchange opportunities are designed to foster true community between the two institutions, strengthen the existing partnerships, and build lasting foundations for long-term

success. Project evaluation efforts are rooted in the concept of "self-authorship," which is the ability to define one's own beliefs and create one's own identity. We will examine how changes to their environment (e.g., moving from natural science to data science, moving from an HBCU to a PWI) impact students' understanding of themselves and their professional identities.

Introduction and Problem Statement

By many estimates, more than a quintillion bytes of data are created every day [1]. This is the equivalent of generating 2-3 hours of high quality video (about 130 megabytes) for every person on earth – every single day – and the pace of data generation continues to increase. Data scientists are trained to use computational, mathematical and statistical methods to understand and analyze these vast datasets, identifying the important information and using it to help guide decisions in fields as diverse as psychology [2], aviation [3], supply chain management [4], and educational methods in STEM (science, technology, engineering, math) [5]. However, there are not nearly enough data scientists to meet the growing demand: more than 2.7 million job openings for data scientists were expected by 2020 [6], and the U.S. Bureau of Labor Statistics estimates a 15% increase in demand by 2029 [7].

These massive datasets have tremendous potential, yet a lack of access and knowledge has created a divide between those who can leverage big data to understand and influence the world around them, and those who cannot. These disparities occur not just outside scientific disciplines, but also within STEM fields due to differences in preparation and experience that create barriers and impede access to data science in certain communities. For example, college students earning degrees in the biological sciences often graduate with limited training in mathematics and statistics and little or no exposure to computer programming. Without sufficient background in mathematical, statistical, and computational approaches, these graduates cannot leverage advances in big data to pursue innovations in biology and other life sciences. Furthermore, there are deep, historical barriers to participation in STEM based on race, gender, disability, and socio-economic status [8]–[12]. These barriers have a substantial social and economic costs, and working to increase access to data science training can both broaden participation in STEM and begin to address some of these historical inequities.

Research supports the importance of broad representation in data science: diverse teams are more likely to identify and avoid biases, resulting in data sets and algorithms that are less likely to produce biased predictions [13]–[15]. Algorithms with encoded prejudices produce biased results that can negatively influence marginalized populations. One such example is the Correctional Offender Management Profiling for Alternate Sanctions (COMPAS) algorithm, which was used by the U.S. court system to make decisions about bail and sentencing [16]. The COMPAS algorithms were inherently biased and caused significant harm: they incorrectly predicted that recidivism was twice as likely for black defendants, while significantly under-estimating the likelihood that white defendants would re-offend [17], [18].

In response to this growing need for more – and more diverse – data scientists, educational institutions have developed a variety of majors, minors, and certificate programs to make data science training more accessible and prepare students for data-driven careers in STEM, business, healthcare, social sciences, and beyond [19]–[22]. The National Science Foundation (NSF) is

also working to build a "Data Science Corps" through collaborative efforts between academia, industry, and communities. The NSF has invested in a range of research projects that are trying to identify both local solutions and broader best practices for "supporting programs designed to help students add data science expertise to their skillset while also helping community-based and nonprofit organizations take advantage of the data they possess" [23].

Vision and Goals: Partners for Change

In an effort to broaden the data science workforce, a Predominantly White Institution (PWI) and an Historically Black College or University (HBCU) are partnering to develop new opportunities for students and faculty at both institutions. The HBCU is Spelman College, an institution nationally recognized for excellence in supporting minority students; it offers robust undergraduate programs in biology, physics, and math – but currently has no formal data science program. The PWI is Michigan State University (MSU), a large, research-intensive institution with experience developing data science programs for undergraduates, graduate students, and working professionals seeking to upskill. Faculty from both institutions have been collaborating for about two years to understand the unique strengths and needs of their home campus and how they can combine efforts to broaden participation and diversify the data science workforce. Five years of funding have been secured from the National Science Foundation and the Arthur P. Sloan Foundation to support this project, which includes four complementary efforts to increasing access to data science across STEM:

1. **3+2 BS+MS:** we are creating a five-year pathway that will allow undergraduates to earn a Bachelor of Science (BS) in a STEM field (e.g., Mathematics, Physics, Biology) from Spelman, and a Master of Science (MS) in Data Science from MSU.
2. **DS at HBCU:** we are developing and implementing new curriculum at Spelman to expose students and faculty to data science at multiple levels, from individual lessons incorporated within existing STEM courses to opportunities to pursue a major or minor in Data Science (DS).
3. **Faculty Training:** we are developing a faculty exchange and training program that will prepare Spelman faculty to integrate data science into their own teaching and research, while simultaneously preparing faculty at MSU to effectively recruit, support and mentor a more diverse student population.
4. **Building a Consortium:** we have begun conversations with a consortium of HBCUs interested in learning from and adapting these efforts to expand access to data science; while this is a longer-term goal, we are committed to collaborating with these institutions to bring similar opportunities in data science to their students and faculty.

These efforts will equip students and faculty in STEM with the computational and mathematical skills necessary to leverage data within their own domains. Many employers are seeking disciplinary experts who are also data science-literate, so this approach gives students a competitive edge in the job market. Demographics vary significantly between different STEM disciplines, with greater diversity in some areas (like biology and health science) than in others (like engineering and computer science). Thus, our efforts to train students and faculty in other STEM fields can help to increase diversity in data science, which is currently 26% female, 7.8%

Hispanic/Latino, and 4% Black [24], [25]. This approach can also help improve the diversity of STEM graduate programs, which have long faced challenges in recruiting and retaining minority students due to a variety of personal, social, and institutional factors that can undermine persistence [26].

3+2 BS+MS: Enhancing STEM with Data Science

Creating a 3+2 program will allow students to complete a bachelor's degree in STEM from Spelman and a master's degree in Data Science from MSU within a five year timeframe. As a starting point, we are working with the Spelman Biology, Mathematics, and Physics programs to identify students whose interests and career goals are well-aligned with the 3+2 program; however, this opportunity is open to any interested STEM student at Spelman. We are seeking to identify cohorts of students (3-5 per year) who will move through the program together and form a natural community of support during the transitions between majors and campuses.

MSU has extensive experience in developing and implementing successful data science training for different audiences and in different formats. For nearly a decade, faculty at MSU have been providing focused training in mathematical and computational skills designed for specific audiences and purposes. One highly effective approach has been developing a catalog of modular, one credit, "short courses" that can be completed in a single month. Each course focuses on specific skills, such as: Programming Foundations; Statistical Analysis and Visualization of Biological Data; Intro to Data Handling (Unix and Python); Metagenomics; Classical Sequence Analysis; and Gaps, Errors and Missteps in Statistical Data Analysis [27]. Students can take just one short course to build a specific skill set, or combine multiple short courses to create a full semester (3 months) of training.

Building on the success of this approach, students in the 3+2 program will be able to select from a catalog of short courses designed to complement their STEM education. For example, students majoring in Mathematics might select short courses focusing on computer programming and machine learning, while students majoring in Biology might opt for short courses focusing on foundational skills in mathematics and statistics. These courses will be co-developed and co-taught by faculty at Spelman and MSU, taking advantage of remote learning methods to allow students on both campuses to enroll simultaneously in the same class. In some cases, these courses may be fully online experiences, while at other times students might gather with each other on their campuses to work with a faculty member who is teaching remotely, with hands-on support from teaching assistants in each classroom.

MSU faculty have also launched a variety of Data Science programs in the last five years, including two different bachelor's degrees; an undergraduate minor; and a master's degree. The new MS in Data Science program is designed for students with a STEM background who want to add a credential in data science. Students at MSU may opt to pursue this MS through dual-enrollment as an undergraduate (similar to the 3+2 program we are building with Spelman), or as an entirely separate program completed after earning the bachelor's degree, with or without intervening years of work experience. Michigan State's MS in Data Science is a 30 credit program designed to be completed over two years, full-time. There are 6 core courses covering essential knowledge for a data scientist (18 credits total); 3 cognate courses (9 credits total); and

one capstone course (3 credits). The curriculum is designed with the assumption that incoming students have some background in Python programming, both single- and multi-variable calculus, and applied linear algebra.

Spelman students pursuing the MS in Data Science at MSU through the 3+2 program we are building will complete the same requirements, with two notable differences. First, the 9 credits of cognate coursework are completed through a series of hands-on practicum studies in faculty-mentored research related to the student's other interests in STEM. Second, a total of 9 credits from the MS program will also be applied to the requirements of the BS in STEM from Spelman. This "double counting" allows students to complete both degrees within a five-year timeframe while ensuring that they receive appropriate training in both their STEM major and in Data Science.

The typical five-year pathway for students in the 3+2 BS+MS program is organized so that students spend the first three years physically at Spelman, and the final two years on the MSU campus. In **years 1 and 2**, students will complete the standard undergraduate coursework for their major at Spelman. In **year 3,** students will finish their disciplinary coursework and take 3 credits of "short courses" co-taught by Spelman and MSU faculty. These short courses (each 1 credit, lasting 1 month) help prepare students for graduate studies in data science; more specifically, students will take: Introduction to Multidimensional Calculus for Data Science; Introduction to Data Structures and Algorithms; and Introduction to Applied Linear Algebra.

Years 4 and 5 are spent at MSU, where students will complete graduate coursework and engage in hands-on learning experiences designed to build skills in data science as it applies to other STEM domains. The typical course of study at MSU includes:

- **Fall 1:** one graduate class, Practicum Rotation
- **Spring 1:** one graduate class, Practicum Project
- **Summer:** Practicum Project (full-time research, no classes)
- **Fall 2:** two graduate classes, Practicum Project
- **Spring 2:** two graduate classes, Capstone

The Practicum Rotation allows students to spend one month each in three different research labs/groups, with the goal of identifying the one research lab/group where students want to complete their Practicum Project. This three-semester Practicum Project offers students the opportunity to work collaboratively with faculty and other students as they gain deep experience in the challenges of working with messy, real-world data in an application area related to their undergraduate studies in STEM. The Capstone is a graduate course that allows students to synthesize what they have learned and prepares them for success after graduation, whether they choose to pursue a career or continue their education in a PhD program.

As part of their graduate coursework, students will take classes from three different departments at MSU, which allows them to build connections with faculty and students from across campus. These units include the Department of Computational Mathematics, Science, and Engineering (CMSE), the Department of Statistics and Probability (STT), and the Department of Computer

Science and Engineering (CSE). Each contributing department teaches two of the six core courses in the Data Science graduate program, which are:

- **STT 810: Mathematical Statistics for Data Scientists.** A streamlined introduction to probability and statistical theory, with an emphasis on the tools that are most relevant to data scientists.
- **STT 811: Applied Statistical Modeling for Data Scientists.** An introduction to major applied statistical models including linear and generalized linear regression, logistic regression, variable selection, and basics of experimental design.
- **CMSE 830: Foundations of Data Science.** An introduction to the core mathematical principles that underlie the algorithms and methods used in data science, including applications to practical problems.
- **CMSE 831: Computational Optimization.** Introduction to applications and algorithms for finite-dimensional linear and non-linear optimization problems, including practical applications.
- **CSE 482: Big Data Analysis.** Data collection, storage, preprocessing, and analysis techniques. Programming for large-scale data analysis. Case studies and applications.
- **CSE 881: Data Mining.** Techniques and algorithms for knowledge discovery in databases, from data preprocessing and transformation to model validation and post-processing. Application of data mining to various domains.

The CMSE and STT courses were designed specifically for the MS degree, while the CSE courses have been adapted for this degree. Collectively, these courses provide the mathematical, statistical, and computational knowledge and practice necessary for students from a wide variety of STEM backgrounds to become effective data scientists.

DS at Spelman: Expanding Options in Data Science

An early step was to assess the existing curricula at Spelman and MSU and identify areas where additional training and supports might be needed for students to successfully transition from a Bachelor's program in Natural Science to a Master's program in Data Science. For example, at Spelman students majoring in Biology, Physics and Mathematics develop a strong foundation in computer programming (Python) as part of their existing curriculum. However, success in Data Science requires additional preparation in applied linear algebra and code optimization that are not typically part of natural science degrees (at Spelman or elsewhere). Towards this goal, faculty at both institutions are collaborating to adapt successful courses from MSU to provide more opportunities for Spelman students to build skills for success in Data Science.

At MSU, faculty have completely redesigned how linear algebra is taught at the undergraduate level, creating an "Applied Linear Algebra" course (MTH/CMSE 314) that focuses on applications and solving large scale problems using computers. The course uses a flipped classroom model, where students watch lecture videos and engage with course materials in advance, and then spend class time working with instructors to solve problems, both individually and in small groups. This course has been taught and refined by a number of different instructors

at MSU over several semesters, and all of the course materials have been developed as Open Educational Resources that are available to students at no cost. As part of this project, we are working to integrate this course into Spelman's curriculum and determining the logistics of offering this class simultaneously on both campuses. Students will have the same lectures and course materials, accompanied by opportunities to work in small groups both on their home campus and virtually with students and faculty from the partner institution.

To provide more specific background knowledge that is important for data scientists, we are also developing new "short courses" that allow students to build skills in areas not typically included in their primary degree programs. For example, a new short course in "Code Optimization for Big Data Analysis" will offer additional preparation in algorithm optimization, data structures, algorithm complexity (Big-O), and basic graph theory (tree structures) as they relate to data science. Another new offering ("Essentials of Multi-Variable Calculus for Data Science") will support students who are not majoring in fields where a full semester of multi-variable calculus is required, but who need to develop some key, applied skills for data science (like using numerical methods for gradients and solving partial differential equations). By developing targeted short courses, we are able to bridge gaps between the requirements of existing undergraduate majors and the skills needed to succeed in the Master's program without increasing graduation requirements or extending the time to degree.

In addition to serving the needs of students pursuing the 3+2 BS+MS program, these new courses will be available to other students at Spelman who are interested in exploring data science. Students might choose to gain a specific skill through a short course; take a sequence of classes to dive a little deeper; or add a minor or major in data science to their undergraduate degree. Adding a data science minor to the Spelman catalog will create opportunities for students from all disciplines to develop data literacy and to understand how data science and its tools are used in their field of study. The minor is interdisciplinary, modular, and accessible to students from a range of majors with different levels of math and computing skills. More specifically, students opting for the minor will complete six courses: one course in each of four core competency areas, and two elective courses. The core competencies are:

1. **Data Science Literacy.** An introduction to data science, data ethics, and the use of data for public benefit or detriment, with emphasis on how data are used to represent marginalized and/or minority communities.
2. **Mathematics in Data Science.** Key lessons in statistics, linear algebra, and other math concepts essential for data science.
3. **Algorithmic Thinking.** Computer programming in Python or R.
4. **Data Curation and Analysis.** Data gathering, data visualization, data analysis, and software tools used by data scientists.

To round out the minor, students will choose two electives. This might include courses from their discipline that emphasize the use of data, or courses elsewhere on campus that deepen students' skills in data science. This flexibility gives students opportunities to design the minor that best aligns with their educational and career goals.

Faculty Training: Sharing Knowledge, Building Community

While the two partner institutions share a vision of expanding access to data science education, there are significant differences in their campuses. Spelman College is a comprehensive four-year institution serving approximately 2,000 undergraduates, with 95% of these students identifying as Black. Faculty at Spelman have significant teaching loads as well as a commitment to mentoring students and building community on a small campus where the student to faculty ratio is 10:1. The research and scholarship of Spelman faculty is an extension of their commitments to teaching and mentoring undergraduates, and the institution is recognized nationally for excellence in supporting minority students [28].

In contrast, Michigan State University is a large public research institution, with nearly 40,000 undergraduates and more than 11,000 graduate and professional students, over 70% of whom identify as White. The most recent research expenditure totals (from 2019) topped $725 million across campus, including more than $330 million in federally funded projects [29]. In response to the growing demand for data science, MSU has made significant investments in the last decade to build new departments, undergraduate majors and minors, and graduate degrees and certificate programs, including hiring more than 20 new faculty with research expertise in data science [30].

In addition to adapting and expanding data science curricula, this partnership includes opportunities for faculty exchange and training programs, both physical and remote. For example, a Spelman faculty member from Biology might take and/or co-teach a "short course" focused on developing foundational mathematical and computational skills, then work with MSU faculty to design lessons that introduce data science into STEM classes on both campuses. Another Spelman faculty member might opt to spend a few weeks of the summer term living and learning at MSU, working alongside faculty and researchers to gain new skills and access additional resources to further their own scholarship. The external funding secured for this partnership will make these types of faculty training experiences possible through teaching release or by providing additional resources that help create time for Spelman faculty to learn about data science and integrate these skills into their classrooms and labs.

For MSU, this partnership offers opportunities to learn from a national leader in recruiting and graduating minority students. Spelman has more than 140 years of experience in educating minority women, and demonstrated success in holistic admission, in creating a culture of support and community across campus, and in maintaining a national network of engaged alumni. Spelman has recently been designated by the Department of Defense as a Center of Excellence for Minority Women in STEM (COE-MWS) [28]. The goals of the COE are to strengthen faculty and student research and engagement in emerging STEM fields; to develop students who are interested in pursuing graduate degrees in STEM; to host conversations that showcase cutting-edge, interdisciplinary research that is conducted by prominent women of color in the STEM disciplines; and to create and disseminate best practices for the preparation and advancement of minority women in STEM. The COE is just one example of how Spelman excels at identifying, implementing, and disseminating best practices on educating and advancing minority students. These best practices will be shared with MSU to assist in efforts to support

students in the 3+2 program once they transition to graduate study, and in broader efforts at MSU to recruit and retain a diverse student population.

Understanding and embracing culture is central to the success of this collaboration, which spans two very different institutions and draws faculty and students from many disciplines. Each individual brings their own cultural heritage and unique perspective to the project, while the learning and research activities take place across distinct institutional, disciplinary, and community contexts. Faculty from both campuses are working collaboratively to identify and address barriers that impact student retention, support for minority graduate students, and support for minority faculty seeking tenure. We are also collaborating on activities that will broaden the conversation about data science in STEM, like a joint seminar series featuring scholars from diverse backgrounds and data science projects that seek to understand and address social inequities.

Another important goal is ensuring the success of students in the 3+2 program, who will be navigating two very different campus cultures. Faculty from both institutions are collaborating to design activities and experiences that help students situate their academic and professional interests within the context of the 3+2 program, transition successfully between degrees and institutions, and prepare for careers in data science. Key activities include individual and small group mentoring at both institutions; training in the responsible conduct of research; and professional development programs to build students' communication, teamwork, and leadership skills, and to promote diversity, equity, and inclusion.

Building a Consortium: Broadening Participation in Data Science

One longer-term goal of this collaboration is to bring data science to other campus communities, particularly other small, minority-serving institutions that may lack the resources to build data science programs from scratch. We have begun conversations with a consortium of HBCUs interested in learning from and adapting these efforts to expand access to data science. Spelman College has a history of successfully developing collaborations with other institutions, including creating opportunities for students to cross campus boundaries to take classes, conduct research, earn dual degrees, and participate in extracurricular activities. For example, Morehouse College (serving Black men) is considered a sibling institution with many shared classes and activities. The first cohort of students recruited to the MS in Data Science through this partnership is expected to include a Morehouse alumnus, in addition to several Spelman graduates, which demonstrates the potential for leveraging existing partnerships and growing new opportunities.

Both Spelman College and MSU have experience developing Memoranda of Understanding (MOUs) for academic collaborations like 3+2 programs. For example, since 1969 Spelman has offered its students a dual-degree program in engineering in collaboration with 14 partnering institutions. Other examples of successful academic collaborations for Spelman College include 3+2 BS+MS programs combining a BS in Biology with a MS in Neuroscience, or a BS in Health Science with a MS in Public Health. MSU has MOUs for dual-degree graduate programs in engineering disciplines with a number of international institutions, as well as a 3+2 program with Medgar Evers College (an HBCU) and graduate programs in Forestry at MSU. Our plans to build a consortium offering undergraduate and graduate training in data science beyond the

confines of this single HBCU+PWI collaboration build on the success of these previous experiences. While we anticipate that it will take a few years to realize these broader goals, we see them as a way to continue expanding access to data science far beyond the five year timeline for building the 3+2 program.

Project Evaluation Plans

This project allows us to intentionally welcome more minority students into Data Science and to expand their opportunities in STEM fields more broadly. In order to understand whether the classes, programs, exchanges, and other activities we are developing have the desired impacts, we will gather data from participants through surveys and interviews. Most of our evaluation plans center on understanding the impact on and perspectives of minority students. Our research and evaluation plans prioritize the perspective, voices, and lived experiences of these students and draws upon qualitative methods (specifically, narrative theory), to see participants as co-creators of new knowledge around the impacts of this program.

We are focused on how participants come to create and clarify for themselves an identity as data scientists and what factors may shift if and how a person comes to see themselves as a "professional" [31], [32]. We root this work in the concept of "self-authorship," or "the internal capacity to define one's beliefs, identity and social relations" [32]. The process of self-authorship is a journey, from questioning one's beliefs, to clarifying beliefs, to authoring a set of individual beliefs: this journey is one many college students go through as they move through situations and negotiate new information, contexts, and relationships that challenge their foundational beliefs entering college, to emerging with a clarified, "self-authored" sense of self [33], [34].

For the participants in this program, there are several different changes to their environment that are likely to influence their cognitive, interpersonal, and intrapersonal beliefs around their identities as data scientists. Moving from Spelman College to MSU; moving from a learning experience dominated by minority students to one where they are a numerical minority; moving from a STEM major to a Data Science major: these types of environmental shifts can potentially prompt powerful considerations about student/professional identity. We are planning a series of interviews with each participant that will allow us to examine the evolution of attitudes and beliefs that shape one's perception of self within the context of their work:

- **Entry Program Interview:** To ask participants how they "see" themselves as students, identify what forms of supports they find most beneficial in their development as scholars, and how they describe the integration of their personal and professional values as they imagine their professional selves.
- **Mid-Point Interview:** To ask participants if/how their own understanding of self has evolved in the context of their academic work, identify if/how the program supports are affecting their experience in real-time, and if/how they are learning what it means to be a professional in their field.
- **Exit Program Interview:** Ask each participant to describe their professional identity within the context of their work, identify which program supports they believe at this point to be most useful in helping support their experience, and how they gained insights as they move into professional experiences.

Throughout the process of these interviews, we will use the voices of participants to track, over time, how their understanding of themselves within the context of their work may be shaped. We anticipate that these insights will contribute to the scholarship on how students can best be supported to embrace a professional identity.

Initial Efforts and Ongoing Plans

The team that developed this proposal has been working together for nearly two years, although external funding became available only in the fall of 2021. Within the first six months, the project team has already recruited the first cohort of students from Spelman (and Morehouse) to earn their MS in Data Science from MSU starting in the Summer of 2022. This first cohort of students is being recruited from among those STEM undergraduates expecting to graduate in May, 2022, which means they will not have time to complete preliminary data science coursework at Spelman as part of their undergraduate studies. Instead, they will start with a "bridging" summer at MSU. The students will live and learn in community along with other STEM undergraduate and graduate students, and will work as a cohort to strengthen their foundation in mathematical and computational skills and prepare for graduate studies.

To give students time to make a successful transition to their new campus, they will spend the first two months of the summer program living on campus, with free room and board and a living stipend. A number of activities are being planned to help students explore campus and the surrounding communities, with a focus on helping students identify where they want to live during their graduate studies. There are a number of on- and off-campus options, and with no housing expenses for the first two months students should be able to save part of their living stipend to help with start-up costs for apartment life. Once students pick a place to call home, we will assist with moving and settling in during August, with the fall semester starting at the end of that month.

All students admitted to the 3+2 program will receive full funding for their MS program, including a tuition waiver, health insurance, and a living stipend. This first cohort of students will start after completing their bachelor's degree, thus extending the time to MS to almost 6 years, however they will have a "full ride" for the additional graduate training. As we begin to recruit and prepare Spelman students earlier in their STEM programs, we expect that the total costs for students in the 3+2 program will be less than what they might otherwise pay for their bachelor's degree, as they will only be responsible for costs in their first 3 years.

In addition to recruiting and preparing to support our first cohort of Data Science MS students, we are working to build community between faculty and students on both campuses. One project that is already underway is a shared seminar series featuring external speakers from diverse backgrounds sharing their data science scholarship. Due to the ongoing pandemic, the seminars in winter/spring 2022 were fully virtual, with a presentation from the speaker and Q&A from the public. We also organized a separate follow up session for the speaker to meet with just students, which was a well-received opportunity for students to ask questions and learn more about the speaker's education and career pathways, and their experiences in data science.

Acknowledgements

This material is based upon work supported by the National Science Foundation under Grant No. 2123260. Any opinions, findings, and conclusions or recommendations expressed in this material are those of the author(s) and do not necessarily reflect the views of the National Science Foundation. This project has received funding from the Alfred P. Sloan Foundation under grant number G-2021-16976.

References

[1] K. Domdouzis, P. Lake, and P. Crowther, "Big Data," in *Concise Guide to Databases: A Practical Introduction*, K. Domdouzis, P. Lake, and P. Crowther, Eds. Cham: Springer International Publishing, 2021, pp. 141–163. doi: 10.1007/978-3-030-42224-0_6.

[2] L. Qiu, S. H. M. Chan, and D. Chan, "Big data in social and psychological science: theoretical and methodological issues," *J Comput Soc Sc*, vol. 1, no. 1, pp. 59–66, Jan. 2018, doi: 10.1007/s42001-017-0013-6.

[3] G. Gui, Z. Zhou, J. Wang, F. Liu, and J. Sun, "Machine Learning Aided Air Traffic Flow Analysis Based on Aviation Big Data," *IEEE Transactions on Vehicular Technology*, vol. 69, no. 5, pp. 4817–4826, May 2020, doi: 10.1109/TVT.2020.2981959.

[4] B. Roßmann, A. Canzaniello, H. von der Gracht, and E. Hartmann, "The future and social impact of Big Data Analytics in Supply Chain Management: Results from a Delphi study," *Technological Forecasting and Social Change*, vol. 130, pp. 135–149, May 2018, doi: 10.1016/j.techfore.2017.10.005.

[5] F. Benita, D. Virupaksha, E. Wilhelm, and B. Tunçer, "A smart learning ecosystem design for delivering Data-driven Thinking in STEM education," *Smart Learn. Environ.*, vol. 8, no. 1, p. 11, Jul. 2021, doi: 10.1186/s40561-021-00153-y.

[6] S. Miller and D. Hughes, "The Quant Crunch: How the Demand for Data Science Skills is Disrupting the Job Market," Burning Glass Technologies, BHEF, and IBM, 2017. [Online]. Available: https://www.bhef.com/sites/default/files/bhef_2017_quant_crunch.pdf

[7] "Occupational projections and worker characteristics : U.S. Bureau of Labor Statistics." https://www.bls.gov/emp/tables/occupational-projections-and-characteristics.htm (accessed Jan. 28, 2022).

[8] M. Renken, J. Scott, P. Enderle, and S. Cohen, "'It's not a deaf thing, it's not a black thing; it's a deaf black thing': a study of the intersection of adolescents' deaf, race, and STEM identities," *Cultural Studies of Science Education*, vol. 16, pp. 1105–1136, Dec. 2021, doi: 10.1007/s11422-021-10023-1.

[9] S. Vakil and R. Ayers, "The racial politics of STEM education in the USA: interrogations and explorations," *Race Ethnicity and Education*, vol. 22, no. 4, pp. 449–458, Jul. 2019, doi: 10.1080/13613324.2019.1592831.

[10] M. A. Wilson, M. A. Odem, T. Walters, A. L. DePass, and A. J. Bean, "A Model for Holistic Review in Graduate Admissions That Decouples the GRE from Race, Ethnicity, and Gender," *LSE*, vol. 18, no. 1, p. ar7, Mar. 2019, doi: 10.1187/cbe.18-06-0103.

[11] "(Un)Hidden Figures: A Synthesis of Research Examining the Intersectional Experiences of Black Women and Girls in STEM Education - Danyelle T. Ireland, Kimberley Edelin Freeman, Cynthia E. Winston-Proctor, Kendra D. DeLaine, Stacey McDonald Lowe, Kamilah M. Woodson, 2018." https://journals.sagepub.com/doi/10.3102/0091732X18759072 (accessed Jan. 28, 2022).
[12] "Not All Disadvantages Are Equal: Racial/Ethnic Minority Students Have Largest Disadvantage Among Demographic Groups in Both STEM and Non-STEM GPA - Kyle M. Whitcomb, Sonja Cwik, Chandralekha Singh, 2021." https://journals.sagepub.com/doi/full/10.1177/23328584211059823 (accessed Jan. 28, 2022).
[13] B. Cowgill, F. Dell'Acqua, S. Deng, D. Hsu, N. Verma, and A. Chaintreau, "Biased Programmers? Or Biased Data? A Field Experiment in Operationalizing AI Ethics," *arXiv:2012.02394 [cs, econ, q-fin]*, Dec. 2020, Accessed: Jan. 29, 2022. [Online]. Available: http://arxiv.org/abs/2012.02394
[14] S. Bakken, "Replication studies and diversity, equity, and inclusion strategies are critical to advance the impact of biomedical and health informatics," *J Am Med Inform Assoc*, vol. 28, no. 9, pp. 1813–1814, Aug. 2021, doi: 10.1093/jamia/ocab168.
[15] M. Grossmann, *How Social Science Got Better: Overcoming Bias with More Evidence, Diversity, and Self-Reflection*. New York: Oxford University Press, 2021. doi: 10.1093/oso/9780197518977.001.0001.
[16] "Compas," *Wikipedia*. Jan. 01, 2022. Accessed: Jan. 29, 2022. [Online]. Available: https://en.wikipedia.org/w/index.php?title=Compas&oldid=1063141137
[17] "How We Analyzed the COMPAS Recidivism Algorithm — ProPublica." https://www.propublica.org/article/how-we-analyzed-the-compas-recidivism-algorithm (accessed Jan. 29, 2022).
[18] J. A. Mattu Jeff Larson,Lauren Kirchner,Surya, "Machine Bias," *ProPublica*. https://www.propublica.org/article/machine-bias-risk-assessments-in-criminal-sentencing (accessed Jan. 29, 2022).
[19] "Data Science Careers, Jobs, Salaries | Data Scientists." https://www.calu.edu/academics/undergraduate/bachelors/data-science/jobs-career-salaries.aspx (accessed Jan. 29, 2022).
[20] "Data Science Careers | University of Wisconsin Data Science." https://datasciencedegree.wisconsin.edu/data-science/data-science-careers/ (accessed Jan. 29, 2022).
[21] "Data Science Careers," *DiscoverDataScience.org*. https://www.discoverdatascience.org/career-information/ (accessed Jan. 29, 2022).
[22] "11 Data Science Careers Shaping the Future," *Northeastern University Graduate Programs*, Jun. 04, 2020. https://www.northeastern.edu/graduate/blog/data-science-careers-shaping-our-future/ (accessed Jan. 29, 2022).
[23] "Developing the 21st century data science workforce," *Beta site for NSF - National Science Foundation*. https://beta.nsf.gov/science-matters/developing-21st-century-data-science-workforce (accessed Jan. 29, 2022).
[24] "The STEM Gap: Women and Girls in Science, Technology, Engineering and Mathematics," *AAUW : Empowering Women Since 1881*. https://www.aauw.org/resources/research/the-stem-gap/ (accessed Jan. 29, 2022).

[25] Priceonomics, “The Data Science Diversity Gap,” *Forbes*. https://www.forbes.com/sites/priceonomics/2017/09/28/the-data-science-diversity-gap/ (accessed Jan. 29, 2022).
[26] F. Curry and J. DeBoer, “A Systematized Literature Review of the Factors that Predict the Retention of Racially Minoritized Students in STEM Graduate Degree Programs,” in *2020 ASEE Virtual Annual Conference Content Access Proceedings*, Virtual On line, Jun. 2020, p. 34069. doi: 10.18260/1-2--34069.
[27] “Bioinformatics Program.” https://cmse.msu.edu/academics/bioinformatics-program/ (accessed Jan. 29, 2022).
[28] “About the Center of Excellence for Minority Women in STEM | Spelman College.” https://www.spelman.edu/coe-mws/about-us (accessed Jan. 29, 2022).
[29] “Facts & Figures | Office of Research and Innovation.” https://research.msu.edu/facts-figures (accessed Jan. 29, 2022).
[30] “About.” https://cmse.msu.edu/about/ (accessed Jan. 29, 2022).
[31] L. S. Nadelson *et al.*, “Am I a STEM professional? Documenting STEM student professional identity development,” *Studies in Higher Education*, vol. 42, no. 4, pp. 701–720, Apr. 2017, doi: 10.1080/03075079.2015.1070819.
[32] “Making Their Own Way.” https://styluspub.presswarehouse.com/browse/book/9781579220914/Making-Their-Own-Way (accessed Jan. 29, 2022).
[33] J. E. Pizzolato, “Assessing self-authorship,” *New Directions for Teaching and Learning*, vol. 2007, no. 109, pp. 31–42, 2007, doi: 10.1002/tl.263.
[34] J. E. Pizzolato and A. B. Olson, “Exploring the Relationship between the Three Dimensions of Self-Authorship,” *Journal of College Student Development*, vol. 57, no. 4, pp. 411–427, May 2016, doi: 10.1353/csd.2016.0052.