The American Statistician

ISSN: 0003-1305 (Print) 1537-2731 (Online) Journal homepage: www.tandfonline.com/journals/utas20

Near-Peer Mentoring in Data Science: A Plot for Mutual Growth

Chiara Sabatti & Qian Zhao

To cite this article: Chiara Sabatti & Qian Zhao (2025) Near-Peer Mentoring in Data Science: A Plot for Mutual Growth, The American Statistician, 79:4, 529-537, DOI: 10.1080/00031305.2025.2550314

To link to this article: https://doi.org/10.1080/00031305.2025.2550314

View supplementary material

Published online: 10 Oct 2025.

Submit your article to this journal

Article views: 70

View related articles

View Crossmark data

Full Terms & Conditions of access and use can be found at
https://www.tandfonline.com/action/journalInformation?journalCode=utas20

THE AMERICAN STATISTICIAN
2025, VOL. 79, NO. 4, 529–537: Teacher's Corner
https://doi.org/10.1080/00031305.2025.2550314

Taylor & Francis
Taylor & Francis Group

Check for updates

Near-Peer Mentoring in Data Science: A Plot for Mutual Growth

Chiara Sabatti[a] and Qian Zhao[b]

[a]Departments of Biomedical Data Science and of Statistics, Stanford University, Stanford, CA; [b]Department of Mathematics and Statistics, University of Massachusetts Amherst, Amherst, MA

ABSTRACT

Universities have been expanding undergraduate data science programs. Involving graduate students in these new opportunities can foster their growth as data science educators. We describe two programs that employ a near-peer mentoring structure, in which graduate students mentor undergraduates, to (a) strengthen their teaching and mentoring skills and (b) provide research and learning experiences for undergraduates from diverse backgrounds. In the Data Science for Social Good program, undergraduate participants work in teams to tackle a data science project with social impact. Graduate mentors guide project work and provide just-in-time teaching and feedback. The Stanford Mentoring in Data Science course offers training in effective and inclusive mentorship strategies. In an experiential learning framework, enrolled graduate students are paired with undergraduate students from non-R1 schools, whom they mentor through weekly one-on-one remote meetings. In end-of-program surveys, mentors reported growth through both programs. Drawing from these experiences, we developed a self-paced mentor training guide, which engages teaching, mentoring and project management abilities. These initiatives and the shared materials can serve as prototypes of future programs that cultivate mutual growth of both undergraduate and graduate students in a high-touch, inclusive, and encouraging environment.

ARTICLE HISTORY
Received March 2024
Accepted July 2025

KEYWORDS
Data science education;
Experiential learning;
Inclusive mentoring;
Project-based learning

1. Introduction

The education literature documents multiple advantages of training opportunities where graduate students work with undergraduates in mentoring capacities. First, involving graduate students can help decrease the student-to-instructor ratio in many courses (Sorte et al. 2020). Second, the access to mentors increases retention of first generation and under represented minorities (URM) in Science, Technology, Engineering and Medicine (STEM) disciplines (Wilson et al. 2012). Graduate students are *near-peer mentors*: while more advanced in training and education, they are of a similar age to their mentees (Garcia et al. 2021). Near-peer mentoring in particular has been shown to be an effective mechanism with which to foster a sense of belonging and self-identity as a scientist (Zaniewski and Reinholz 2016; Wilson and Grigorian 2019). Third, acting as mentors positively influences the personal, cognitive, and professional growth of graduates student mentors (Dolan and Johnson 2009; Tenenbaum et al. 2014). Opportunities that value interdependence and recognize communal goals may increase some students' cultural alignment with the research institution (Stephens et al. 2012; Boucher et al. 2017; Anderson et al. 2019; Belanger et al. 2020).

The possible benefits of near-peer mentoring resonate particularly well with two challenges we encounter in data science education: the need to increase access to research experience, and to diversify the workforce. First, while both the Guidelines for Assessment and Instruction in Statistics Education (GAISE) Reports (2016) and Curriculum Guidelines for Undergraduate Programs in Statistical Science (2014) recommend research experiences as crucial means of enhancing learning, limited faculty time to mentor can impede the development and expansion of such opportunities (Nolan et al. 2020). Engaging graduate students in mentoring roles can increase universities' capacity to offer these project-based and experience-centered learning (Rodolfa et al. 2019).

Second, while it is well recognized that increasing diversity expands the talent pool, boosts innovation and improves the nation's global leadership in science and technology (National Academy of Sciences, Engineering, and Medicine 2011), students who identify as female or belonging to minority groups remain a lower proportion of doctorates in the field of statistics (National Center for Science and Engineering Statistics 2022). Since mentoring experience increases the sense of belonging, self-efficacy, and science identity for both underrepresented mentors (Trujillo et al. 2015) and mentees (Zaniewski and Reinholz 2016; Wilson and Grigorian 2019), it can potentially improve retention at both undergraduate and graduate levels. The strong job market in data science (Davenport and Patil 2012), which is projected to grow 36% in the decade between 2023 and 2033 (U.S. Bureau of Labor Statistics 2025), offers concrete opportunities for the students, possibly from URM, toward which we direct outreach and retention efforts.

CONTACT Qian Zhao *qianzhao@umass.edu* Department of Mathematics and Statistics, University of Massachusetts Amherst, Amherst, MA.
Supplementary materials for this article are available online. Please go to *www.tandfonline.com/r/TAS*.
© 2025 American Statistical Association

While research, consulting, or capstones are offered in many departments and they have been described in publications (e.g., Legler et al. 2010; Lazar, Reeves, and Franklin 2011; Smucker and Bailer 2015), literature on statistics programs that engage graduate students in the role of mentors is lacking. In this article, we report on two programs developed at a four-year university, relying on a near-peer mentoring structure to provide strong learning experience for both undergraduate and graduate students. In the Data Science for Social Good (DSSG) summer fellowship program, graduate students mentor undergraduate student teams to tackle a data science project with positive social impact over the course of eight weeks. Stanford Mentoring in Data Science (SMDS) is both a class offered to graduate students and an outreach initiative. Mentors enrolled in this class participate in interactive workshops focused on inclusive mentoring as well as engage in one-on-one meeting with undergraduate mentees. Both programs were introduced in response to graduate students' interests, and graduate students continue to actively take part in program design. Faculty instructors developed program goals and structure; they support the graduate mentors, creating a robust learning experience for both undergraduate and graduate participants. We describe our experiences with the two programs (Sections 2 and 3), and share learning materials we have developed for graduate student mentors (Section 4). We hope that this will encourage and support future near-peer mentoring programs, fostering learning and a sense of belonging for both graduate and undergraduate students.

2. Data Science for Social Good

Data Science for Social Good (DSSG) is summer fellowship where student teams work on data science projects with social good objectives. The impetus of DSSG came from a request by a graduate student who participated in the homonymous program at the University of Chicago and, having found the experience transformative, was determined to create some version of it at our university. Motivated by their testimony and willingness to work toward this goal, we ran a pilot experience and then develop the program over four years. Compared to other DSSG programs nationwide (see Dailey et al. (2021), for a review), our program is unique in that its core objective is to provide educational opportunities for both undergraduate participants and graduate mentors in distinct roles.

2.1. Program Goals

The DSSG program serves three groups: undergraduate participants, graduate student mentors in training, and project partners.

Our undergraduate participants are "beginner data scientists," who have sufficient technical competencies, that is, have taken upper undergraduate courses in statistical inference, linear regression, and have experience using statistical software through at least one course, but without experience applying data science to a real world problem, and may not be committed to a career in this space. We set four broad goals for this group: (a) to engage in an authentic data science project that addresses a societal need; (b) to receive training, mentoring, and practical experience in aspects of "data acumen" identified by National Academy of Sciences, Engineering, and Medicine (2018) least reflected in a typical educational curriculum, for example, ethics, communication, and collaboration; (c) to experience the feeling that data scientists are "people like me"; (d) to strengthen professional skills and increase self-efficacy, feeling empowered to make career choices with a sense of purpose and service. We focus on the following questions when selecting participants: Do they have adequate foundations in data science, statistics and programming? Are they excited to apply data science for social good? What research areas and society issues are they passionate about? What are their experiences with working in a diverse team? What are their commitment to diversity and inclusion?

Our mentors in training are graduate students and postdocs, who have strong technical expertise and some training in teaching (e.g., serving as graduate teaching assistants (TA), or through TA workshops), but limited experience of mentoring, managing a research team, or navigating the challenges of interdisciplinary work. By engaging in DSSG, the mentors build competence in the following areas: (a) mentoring undergraduate participants in an "active learning" environment; (b) lead a data science team project; (c) engage with project partners of different background without the scaffolding provided from an instructor. We select mentors through an open application, encouraging responses in particular from students supported by the data science initiative, and enrolled in PhD programs as Statistics, Computer Science, Biomedical Data Science.

Our project partners are often nonprofit organizations who have a research question that can be addressed through data analysis. Faculty from across the university is invited to contribute project ideas and suggest potential partners. Potential partners's expectations regarding scope and impact are framed with the understanding that undergraduate students carry out the majority of the work.

2.2. Roles of Graduate Mentors

At DSSG, graduate mentors in training take the lead in project development and participants' learning. We now outline in some detail the activities they engage in, clarifying how these strengthen their teaching, mentoring, project management, and leadership skills.

Identify project goals and milestones. Mentors and faculty advisors meet with the project partner prior to the program to achieve a broad vision of the needs and potential of the collaboration. Mentors then scope the project, keeping in mind its societal impact, the learning it provides to participants, feasibility given the time frame, and the ease with which it can be communicated. Working with the project partner, mentors define concrete goals, weekly milestones, and clarify the available resources. An example project milestones is provided in the supplementary material. Mentors typically do not curate data beforehand, in order to expose participants to real-world data and encourage them to take a leading role in determining which aspect of the data can be used to address which part of the challenge.

Project onboarding. Mentors develop an onboarding document, which includes background information, related literature, and

technical resources. At the start of the program, participants go through the material, meet with the project partner, and ask clarifying questions. After the meeting, participants frame their overarching goal for the summer and identify their first set of tasks, which often involve identifying and processing relevant data and exploratory data analysis.

Technical training. Mentors provide technical training to all participants, such as how to write a reproducible document in R or how to use cluster computing, as well as topics that arises during project work. An example of training materials that mentors developed to teach SQL database is provided in the supplement. Mentors also facilitate participants' self-directed learning, guiding them to suitable resources. Through these experiences, mentors learn how to teach in an active learning environment, encourage explorations while integrating learning into project work, support a growth mindset, and foster students' confidence.

Facilitate team meetings. During the program, mentors participate to the daily project team meeting. Discussion topics range from brainstorming potential methods to evaluating different approaches, from presenting to a non-technical audience to developing reproducible workflow. Mentors can also use daily team meeting to provide just-in-time teaching (Simkins and Maier 2010; McGee, Stokes, and Nadolsky 2016), for example, correct misconceptions, or pose questions for discussions. This regular interaction provides an opportunity for the mentor to facilitate successful team work, coordinate a team project, and foster a sense of belonging.

Collaborate with project partners. Each week, the project team meets with faculty advisors and project partner to share progress, collect feedback, clarify directions, and identify areas where the DSSG team needs support. Both the mentor and participants learn to collaborate with domain experts who do not have strong data science background, but can articulate relevance and implications of the project.

Build a data science community. Mentors organize activities to introduce participants to the wide range of data science topics and career paths. For example, members of the data science community are invited to describe their career trajectory or projects in a weekly presentation. Mentors also organize social activities such as improvization workshop and trivia challenge to cultivate a sense of community among participants. An example schedule of the DSSG program is included in the supplementary material. Mentors meet weekly with the faculty advisors to discuss mentoring challenges and make plans for program activities.

2.3. Program Outcomes

We have run the DSSG program for seven summers, between 2019 and 2025. Data in this report refers to the first four. The average number of people involved has been 8 participants, 3 mentors and two faculty members. Demographics details on participants and mentors from 2019 to 2022 are available in Figure 1 in the supplementary material and the partner and project details can be gathered from the program web-pages. We achieved our goal of involving underrepresented minorities: about 50% of participants and 70% of mentors identify as female or other gender identity. More than 70% of participants and 60% of mentors are from a minority group as described in Figure 1.

We have also surveyed DSSG mentors on their experience. Mentors who responded to the anonymous online survey ($n = 9$) reported gain in their capacity in multiple dimensions (see Appendix A). For instance, a vast majority are more confident in their ability to employ different teaching strategies, working with students to set academic and professional goals, setting goals and milestones of a data science project, and communicating progress with other participants and project partners. To quote one of the mentors: "Before DSSG, although I have served as a TA for several classes, I have never had experience mentoring junior students. I learned from my experience that different people have different ways to approach projects or respond well to different mentoring styles, which can be vastly different from mine. I also learned a bit about how to structure team work such that students with different strengths can complement or learn from each other."

This comment highlights that the DSSG can be a significant learning experience for mentors. It also shows that mentoring a team project can be an unfamiliar and at times challenging for graduate students. As we continue developing this program and other courses that are centered around undergraduate research projects, we start to recognize the importance to meaningfully include and engage graduate students who may not have prior experience or consider themselves as ideal mentors. With the goal of supporting graduate mentors in mind, we have developed a set of learning materials for DSSG mentors, that can be applicable to mentors of general data science projects. We describe the guide in detail in Section 4 and discuss the second program in the next section.

3. Stanford Mentoring in Data Science

Stanford Mentoring in Data Science (SMDS) doubles as a class offered to graduate students and an outreach initiative. Graduate students enrolled in the class participate in both weekly in-class meetings and a mentorship practicum. SMDS was launched during academic year 2020–2021, when instruction was entirely remote, and the US society had experienced a number of dramatic events. In particular, the protests and reckoning with racism had engaged the graduate students population, who self-organized in many forms of activism. A program aimed at increasing the diversity of the data science workforce aligned well with their interests and offered them an opportunity to channel their desire for action. We describe the program goals for both graduate students and undergraduate participants in Section 3.1 and outline program structure in Section 3.2.

3.1. Program Goals

The ultimate goal of SMDS is to assure that the data science workforce is well connected to society and reflects its diversity. The program works toward this goal with a dual strategy. On the one hand, SMDS recruits undergraduate students from currently underrepresented groups in data science, connecting them to the field. On the other hand, SMDS educates graduate students on values of diversity in a research environment, equipping them

with techniques and practice to mentor individuals from diverse backgrounds.

The undergraduate participants are at the early stages of their exposure to data science. For example, they may have taken an introductory statistics course, or they might be more advanced in a related program but lack hands-on experience. We aim to motivate, support and nurture their interest in data science through mentoring and introduction of career paths and research topics. To reach out to potential mentees, we have contacted community colleges from Carnegie Math Pathways: participating faculty advertises the program to their students and directly refer candidates. The application form is purposely simple, collecting information on demographics, current school, courses completed in math/computer science/statistics, areas of interest, and plans for the future. When selecting participants, we look at questions like: Do they have enough background to engage in some data exploration? Is there room in their plans to increase and sustain an interest in data science? Do they appear to be in need of mentorship?

Graduate students working in the data science space can enroll into a class (BIODS360), offered for credit. They learn (a) to develop a set of tailored goals for the mentoring relationship based on the needs and aspirations of their mentees; (b) to recognize stereotypes and implicit biases that might shape our behavior and learning and to use strategies that overcome these, creating and maintaining an inclusive environment; (c) to articulate the value of a diverse and inclusive team in tackling data science problems, being informed about current composition of data science workforce, with an eye both on historical context and future goals for our field.

3.2. Program Structure

For the duration of the quarter (10 weeks), graduate students enrolled in the course participate weekly to one in-person class and meet one-on-one online with their undergraduate mentees. In-class meetings consist of interactive workshops, reflection on mentoring experience, and community building activities.

During the first in-class meeting, after an introduction of the course structure, graduate students are divided in pairs and go through "the wallet exercise." Designed by the Stanford d.school (see references for links to instruction and facilitator's guide), this workshop task participants to design an "ideal wallet" for their partner. After interviewing their partner, students brainstorm an initial design, which they refine after iterations of feedback, and eventually come up with a prototype. This activity serves multiple purposes: it acts as an icebreaker, helping to establish a climate for safe exploration and sharing; it underscores the importance of listening and empathy during meetings with participants; it frees mentors from the idea of an "ideal" project and encourages them to jointly develop one experience that fulfilled their mentee's needs; it introduces the idea that false starts are possible and sometimes necessary to find the right path.

The first class concludes with the assignment for each graduate student to identify their mentee, respond to the e-mail that the course organizer would send to the pair establishing contact, scheduling a first meeting during the current week. During the first meeting graduate students are tasked to learn as much as possible about their mentee, share some information about themselves to start developing a relationship, and find a mutually agreeable time to meet online each week. We found that letting graduate mentors self-match to undergraduate participants, based on some of the information collected in the application, for example, name, gender, degree, and area of interest, increased mentors' sense of agency without creating specific biases.

During the second in-class meeting, all participants are invited to join in on Zoom. Everyone briefly introduced themselves, with the aid of icebreaker questions (e.g., "if I were an animal I would be a..., because ..."). The instructors of the class then share examples of directions that the mentee and mentors could take in their work together: preparing for internship applications, learning a data science topic, exploring careers, taking a close look at a dataset on a topic of interest etc. While we try to underscore that each pair had total autonomy in deciding their "curriculum," we have found that being presented with a list of possibilities helped them to articulate needs and interests. The assignment for the second one-on-one meeting is for the mentor and participant to do a first round of brainstorming on activities that they might engage in during the program. We suggest keeping this brief, leaving time to reevaluate during the coming weeks. During the rest of the meeting, we invite mentors to share a bit more on their experience as data scientists, describing the course of study they are involved in and the topic of their dissertation.

In the following weeks, we invite a number of campus experts in the area of diversity, equity and inclusion to facilitate workshops. They introduce and discusse concepts as belonging, identity interference, stereotype threat, imposter syndrome, offering pointers to literature (Stephens et al. 2012; Leslie et al. 2015) as well as personal experiences. A list of workshops we have offered are included in the supplementary material. We found these workshops particularly useful for both the program and graduate participants. Engaging with experts provided a strong signal of the importance of these topics and the amount of scholarly work that has been carried out. Speakers brought to the table years of personal experience which grounded the concepts and solutions presented and established warmth and acceptance in discussions. Connecting with campus resources substantially enlarged the reach of the program, as we continued to leverage these in other contexts.

A number of in-class meetings are devoted to mutual sharing of how the mentor-mentee relationships are evolving, and brainstorming around possible roadblocks. The instructors meet one-on-one with undergraduate participants and bring to the meeting common feedback from participants. By hearing from each other, mentors are better able to notice their own growth and identify successes and challenges. They are more motivated to develop solutions knowing that they will benefit others. Sharing and discussing as a group also fosters a sense of community among graduate students from different academic backgrounds.

Toward the end of the program, we invite all participants to join a synchronous class meeting. This is an opportunity for participants to share their growth during this program, reinforce peer connections, and present ways to continue the peer-relationships. We have organized career panels during one of the classes, inviting data science professionals who had attended

the same schools as our participants. Due to the pandemic and available resources, our participants' experience has been entirely online so far, and we are working to organize an in-person gathering for the conclusion of the program.

Graduate mentors have created a collection of resources over the years: pointers to outreach programs and internships; articles on data science careers; suggestions on resume and cover letters; tutorials on statistics topics and software; and a collection of datasets and basic analysis scripts in R and Python that can be used as a starting point of data exploration and analysis. An example project can be found in the supplementary material. These resources help to take the pressure off mentors to search for or create resources, allowing them to focus on mentoring.

3.3. Program Outcomes

We have offered SMDS five times during 2021–2025. A summary of the demographic and education information for participants and mentors in the first three offerings is provided in Figure 2 in the supplementary material. Nearly half of the undergraduate participants are from underrepresented minority groups in science and engineering professions. We surveyed participants' program experience and attitude in end-of-program anonymous online questionnaires. During meetings with their mentor, mentees often spent time on learning data science skills (59% of time on average), career counseling (30% of time on average), and individual projects (30% of time on average). End-of-program surveys show mentees became more interested in data science and in pursuing future internships. Graduate students' feedback is collected via the University official course evaluation system. The program has been reported to increase mentors' interest to engage in activities to enhance diversity in their working environment. We report survey results in Appendix B.

4. A Guide for Data Science for Social Good mentors

Research have shown that mentor training improves mentor's skills and self-efficacy (Stamp, Tan-Wilson, and Silva 2015). Moreover, it is known that socioemotional and culturally relevant mentoring positively impacts mentee's research experiences (Haeger and Fresquez 2016). These findings, together with our intention to include and engage mentors from diverse academic and personal backgrounds, motivated us to develop a self-paced learning guide, *Growing by mentoring: A guide for Data Science of Social Good mentors*, to share our learning in the two programs and provide support to DSSG project mentors. While grounded in DSSG projects, the themes and activities are broadly applicable to graduate student mentors of data science project teams.

Growing by mentoring: A guide for Data Science of Social Good mentors covers three broad areas: (a) what are characteristics of a good mentor, and how to develop them; (b) how to facilitate the progress of a team working on an interdisciplinary data science project; (c) how to set individualized learning goals and approaches to help students reach them. Organized around these themes, the guide introduces relevant concepts and useful strategies. Each idea is made concrete with an "activity": by completing all of these, mentors will be able to carry out most of the tasks they need to design their mentoring experience. Each section is self-contained, such that mentors can pick and choose specific topics of interest. The guide is included in the supplementary material, and we outline main topics and activities in this section. While we are aware of established research mentor training programs (Pfund et al. 2014), this is perhaps the first set of self-paced materials developed specifically for graduate mentors for data science projects.

4.1. Growth as a Mentor

The first section prepares graduate students to become *intentional mentors*, who have visions of for both themselves and their mentees and plan their course of actions accordingly. This section comprises four topics: goal setting, effective mentoring, belonging, and creating an inclusive learning environment.

We introduce mentoring as a type of support intended to advance mentees toward an important goal (Packard 2016). It is a reciprocal relationship where both the mentee and the mentor can experience personal, professional, and intellectual growth from respectful, effective, and sustained interactions (Galbraith 2003). We invite readers to consider the qualities and actions of their own mentors. This reflection serves as an entry point into discussions of attributes and practices of effective mentors. We leverage the triangular model of mentor competence from the book *On being a mentor* (Johnson 2007) and examples from the article *Nature's guide to mentors* (Lee, Dennis, and Campbell 2007). To conclude, we invite mentors to reflect on their own strength and weaknesses, and set their mentoring goals. This aims to cultivate a growth mindset, encouraging a continued development of these skills throughout the program.

The objective of mentoring is to support mentees to make progress toward their goals. We describe the S.M.A.R.T. (Specific, Measurable, Achievable, Relevant, and Time-Bound) goal setting framework, provide a list of data science skills students may want to improve in, adapted from *The data for good growth map* (Dailey et al. 2021) and Curriculum Guidelines for Undergraduate Programs in Statistical Science from American Statistical Association Undergraduate Guidelines Workshop (2014), and a list of potential questions, to clarify the breadth of topics mentors can discuss with their mentees.

The next two sections raise mentors' awareness of challenges minority students can face and strategies they can use to create an inclusive learning environment. We introduce the concept of belonging from Walton and Wilson (2018) and highlight the importance of belonging in a learning environment. We discuss challenges to belonging for underrepresented students, such as identity interference (Slepian and Jacoby-Senghor 2021), stereotype threat (Steele and Aronson 1995) and previous negative experiences. We then offer strategies to enhance belonging, such as through reframing challenges (Walton and Wilson 2018) and cultivating a growth mindset (Dweck 2006). We refer to resources such as Lang (2016) and Arif et al. (2021). Mentors watch online videos where students speak about issues they have faced, and brainstorm how they can implement these strategies in their projects.

4.2. Growth of the Project

Group projects can offer a rich learning experience for undergraduates to analyze complex data, think in context, and choose

appropriate method as advised in Curriculum Guidelines for Undergraduate Programs in Statistical Science (2014). Leading a collaborative data science project can strengthen mentors' leadership and project management skills. This section intends to prepare mentors in four aspects: initiating and sustaining collaboration, setting concrete project milestones, facilitating discussions, and organizing teamwork.

We borrow from research in meaningful collaboration in community-based participatory research (CBPR). In both CBPR and Data Science for Social Good projects, the objective is often to spawn positive changes, the project is often a systematic inquiry, and both the data science team and the project partner are engaged throughout the process. Mentors first reflect on a case study, to help them contextualize the discussion and set realistic expectations. We then offer a list of suggestions, based on Pasick et al. (2010) and Swann and Murray (2020), accompanied by an example of a DSSG or CBPR project. For instance, we suggest mentors to be flexible and open to alternative approaches. As an example, we describe a DSSG project where the team needed to develop a specialized workflow to use highly sensitive data, and they embraced the challenge as an opportunity to learn privacy consideration and writing efficient code (Foster et al. 2021).

In another activity, mentors are invited to identify overarching objectives and create project milestones. We provide questions they can discuss with their project partner, emphasizing the importance to clarify the need to be addressed, to analyze the impact on communities who may be affected, and to draft milestones. We borrow ideas from the book *Ten steps to successful project management* (Russell 2007).

Unlike project managers who assigns tasks to team members, mentor's role is to facilitate students' learning and discovery. Team discussion serves as a medium where students collectively find a way forward, generate and express insights and discoveries. The third section encourages mentors to be more deliberate and purposeful when facilitating discussions. We present two perspectives of discussion: components (questioning, listening and responding, following up), and structures (framing, encouraging deeper answers, broadening perspectives, refocusing, and closing the loop). These materials are adapted from workshops and resources from centers for teaching and learning from multiple institutions.

Finally, we encourage mentors to form a consensus with the team about project organization and team collaboration, for example, how to keep track of each other's progress, and how to organize code. Together, the four sections serve as a starting point as mentors embark on their journey as a project leader and manager.

4.3. Growth as a Teacher

The third section aims to support mentor's growth as a data science educator by providing tools and practices to articulate learning goals, cultivate an active learning experience, and assess students' learning. This section is largely based on two books: *How learning works: Seven research-based principles for smart teaching* (Ambrose et al. 2010) and *The ABCs of how we learn: 26 scientifically proven approaches, how they work, and when to use them* (Schwartz, Tsang, and Blair 2016).

We introduce both the cognitive aspects of learning using the revised Bloom taxonomy (Anderson et al. 2001) and the more holistic perspective of learning using Fink's taxonomy (Fink 2013). Mentors are then invited to refine learning goals they have set for mentees from these perspectives.

Next, we focus on three approaches mentor can use to facilitate learning during project work: (a) cultivating a collaborative learning environment, (b) fostering meta-cognitive skills, and (c) incorporating active learning. Again, the introduction of relevant concepts is accompanied by concrete data science examples, and activities the mentors can leverage in their program. For instance, we introduce meta-cognition as the process of reflecting on and directing one's own thinking (National Research Council 2001). Then, we describe the steps of problem solving as well as questions mentors can ask students to reflect on each of these steps, following the characterizations by Price et al. (2021). Mentors are invited to come up with additional questions and ask them during team discussions. By modeling the metacognitive process and gently nudging students to reflect on their learning and problem solving process, mentors can help students develop habits with long lasting impact. Taken together, this section serves the dual purpose of enhance mentor's skill as an educator and students' skill as a data scientist in training.

5. Discussion

In exit surveys, participants and mentors alike report an overwhelming positive experience in DSSG or SMDS. While we have collected feedback from participants and mentors, we have not run a case-control study, and analysis of the outcomes is additionally complicated by the fact that we are here reporting results of pilot programs, whose structure has evolved over time, in response to the feedback received. In addition, our main goals of long-term involvement in data science and education is impossible to measure in this short time frame. As these programs take a more permanent place in the offerings in our university, it will be possible to engage in more systematic outcome measurements in the future.

By sharing our experience and developed materials, we aim to facilitate the design of related programs that we imagine centered on the unique context and existing infrastructure in each institution. For universities with existing consulting or interdisciplinary study programs, giving some advising and mentoring role to graduate students, with faculties supervisors meeting weekly with undergraduate and graduate students can be a way to start applying a near-peer mentoring structure. For colleges without a graduate student population, upper-undergraduate students who have more research experience might serve the role of near-peer mentors, with additional structured guidance by faculty advisors.

We conclude this article with some personal reflections on the impact of these program on our experiences as a faculty member and as a graduate student. By design, DSSG and SMDS had graduate students and faculty members working side-dy-side in the organization and execution of programs in the educational space. This was an energizing and enriching experience, as both groups engaged outside the traditional roles of instructor and teaching assistant. Graduate students contributed energy and

Table 1. DSSG mentors' self-identified growth.

Topics		Specific aspects	Mean
Teaching Data Science	T1.	Assessing my mentee's prior knowledge and skills in data science.	1.1 (0.2)
	T2.	Communicating data science topics effectively.	1.2 (0.3)
	T3.	Employing different teaching strategies (e.g., worked example, peer feedback).	1.3 (0.2)
	T4.	Encouraging students to reflect on their work.	0.9 (0.3)
	M1.	Connecting with students from diverse backgrounds.	1.1 (0.3)
Mentoring	M2.	Cultivating a growth mindset.	0.9 (0.2)
	M3.	Working with mentees to set academic and professional goals.	1.3 (0.2)
	M4.	Providing constructive feedback.	1.0 (0.2)
	M5.	Actively listening to mentees' feedback and concerns, and incorporating them into my mentoring.	1.3 (0.3)
	P1.	Gauging the impact of a project.	1.2 (0.1)
Project Management	P2.	Setting goals and milestones for a data science project.	1.4 (0.2)
	P3.	Communicating progress with participants and community partners.	1.1 (0.3)
Leadership, Community building	L1.	Facilitating group collaboration.	1.2 (0.1)
	L2.	Motivating and engaging participants.	1.0 (0.2)
	L3.	Building a community of data scientists.	0.8 (0.3)

NOTE: Mentors rated their change of confidence in each category on a Likert scale, where −2 indicates "much less confident", 0 indicates "No change", and +2 indicates "Much more confident". We report the average of mentors' responses and its standard deviation within the parenthesis.

Table 2. Participants and mentors' feedback at the end of SMDS program.

		Questions	Average
Participants	T1.	What portion of the time was spent on learning specific data science skills?	59% (4%)
	T2.	What portion of the time was spent on career counseling (e.g., CV revision, interview preparation, class selection, etc.?)	30% (5%)
	T3.	What portion of the time was spent working on a project?	29% (4%)
	P1.	How has participating in this experience changed your interest in Data Science?	1.6 (0.1)
	P2.	How has your experience in this program changed how likely you are to apply for other internships/mentorships?	1.7 (0.1)
Mentors	M1.	How has participating in this class changed your interest in mentoring?	1.4 (0.1)
	M2.	How has participating in this class changed your willingness to engage in activities aimed at increasing diversity in the groups of people with whom you work?	1.5 (0.1)

NOTE: Participants and mentors rated their change of interest in each question on a Likert scale, where −2 indicates "Definitely less interested", 0 indicates "No change", and +2 indicates "Definitely more interested". We report the average of responses and its standard deviation within the parenthesis.

fearlessness in pursuing projects and opportunities, and were able to articulate and channel the expectations of the participants. Faculty made sure that there was a fertile ground to sustain this growth: they were able to estimate and garner the resources needed, provide a referential framework with which to evaluate efforts and progress, place the activities in a broader educational context, bring in experts and pointing to methods and research areas. This enhanced collaboration improved the cohesion between faculty and students in the same department, and increased the agency and confidence of graduate students.

Working in a university, and having only a finite amount of time and resources, we experience occasionally tensions between what appear to be different goals: educate a new generation, transmitting the knowledge we have and cultivating in them the aspiration to keep increasing it; engage in research, pushing the boundaries of our understanding; deploy this knowledge in practical applications, impacting the society around us. Being involved in these programs was so enjoyable partly because multiple of these goals aligned: we had the opportunity to connect with many different reality on campus, learn, educate, and make an impact—university at its best.

Appendix A. DSSG Mentor Growth

We surveyed DSSG mentors who participated between the years 2019–2022. Each question is framed as following: "How does DSSG change your confidence in the following aspects of teaching data science (mentoring/project management/leadership and community building)?" Mentors are asked to rate on a Likert scale ("Much less confidence", "Somewhat less confident", "No change", "Much more confident", "Somewhat more confident") with an additional option of "Prefer not to respond". We code the five options using numeric values $(-2, -1, 0, 1, 2)$. Nine mentors took the survey, and almost all of them responded to all of the questions in the range from 0 to 2 (response rate = 70%). We report the mean and the standard deviation of the mean to each aspect in Table 1. We note that the survey was first conducted in June, 2022, which may be a long time away from when some of the mentors participated in DSSG.

Appendix B. SMDS Program Feedback

We surveyed both participants and mentors who participated in SMDS at the end of the program. Altogether, 34 out of 57 undergraduate participants in 2021–2023 have responded to the survey (response rate = 60%). We received 26 responses from the graduate student mentors (response rate = 70%). Because the survey is anonymous, if a mentor participated in the program multiple times, their responses in each year are counted as separate responses. Participants and mentors are asked to rate on a Likert scale ("Definitely less interested", "Slightly less interested", "No change", "Slightly more interested", "Definitely more interested") and we code these five options using numeric values $(-2, -1, 0, 1, 2)$. All of participants and mentors rated all of the questions in the range from 0 to 2. We report the mean and the standard deviation of the mean to each aspect in Table 2.

Supplementary Materials

DSSG program materials Descriptions of activities; example project milestone; example onboarding material; example of learning material.

SMDS program materials List of SMDS workshop topics; example of analysis scripts.

DSSG and SMDS demographics Demographic information of program participants.

"Growing by mentoring: A guide for Data Science for Social Good mentors" A self-paced learning material for DSSG mentors.

Acknowledgments

The value of experiences we described stems from the people who engaged in them: we are thankful to all participants, mentors, faculty and staff who dedicated time and energy to pilot and develop these programs. We thank Ben Stenhoug for his vision to create a Data Science for Social Good program at Stanford and Ben, Emily Flynn and Michael Sklar for agreeing to work as mentors in the first edition, very much pulling ourselves up from our bootstraps, and making a blue print for years to come. Thanks to Balasubramanian Narasihman for having shouldered major responsibilities for this program for seven years, committing all his summers and his good spirits, and being ready to join into the SMDS adventure in unflappable manner. Saara Khan and Jonatan Feiber, from the d.School, have enriched our experience with workshops and masterful handling of interactions over Zoom. Lucy Bernholz provided a much needed connection with the humanities and social science camp on campus. Connor Doherty coordinated many students contribution to SMDS. Miranda Stratton, Marcella Anthony, Joseph Brown, Zandra Jordan volunteered their time and expertises to help us grow our inclusivity muscles. Elizabeth Munoz, Pornprang Plangsrisakul and Natalie LaMariana kept us, and a lot of people outside our university, on track. Bernadine Chuck Fong connected us with an incredible number of partners, starting from the Carnegy Math Pathways, where Dan Ray has been very helpful.

Disclosure Statement

The authors report there are no competing interests to declare.

Funding

Support from the National Science Foundation (NSF 1934578, DMS-2210392) and Stanford Data Science made all of this possible.

ORCID

Qian Zhao http://orcid.org/0000-0002-2287-4871

References

Ambrose, S. A., Bridges, M. W., DiPietro, M., Lovett, M. C., and Norman, M. K. (2010), *How Learning Works: Seven Research-based Principles for Smart Teaching* (Vol. 1), San Francisco, CA: Jossey-Bass. [534]

American Statistical Association Undergraduate Guidelines Workshop. (2014), "Curriculum Guidelines for Undergraduate Programs in Statistical Science," available at *https://www.amstat.org/docs/default-source/amstat-documents/edu-guidelines2014-11-15.pdf*. [529,533,534]

Anderson, L., Krathwohl, D., Airasian, P., Cruikshank, K., Mayer, R., Pintrich, P., Raths, J., and Wittrock, M. (2001), *A Taxonomy for Learning, Teaching, and Assessing: A Revision of Bloom's Taxonomy of Educational Objectives*, New York: Longman. [534]

Anderson, M. K., Anderson, R. J., Tenenbaum, L. S., Kuehn, E. D., Brown, H. K., Ramadorai, S. B., and Yourick, D. L. (2019), "The Benefits of a Near-Peer Mentoring Experience on STEM Persistence in Education and Careers: A 2004–2015 Study," *The Journal of STEM Outreach*, 1, 1–11. [529]

Arif, S., Massey, M. D. B., Klinard, N., Charbonneau, J., Jabre, L., Martins, A. B., Gaitor, D., Kirton, R., Albury, C., and Nanglu, K. (2021), "Ten Simple Rules for Supporting Historically Underrepresented Students in Science," *PLOS Computational Biology*, 17, e1009313. [533]

Belanger, A. L., Joshi, M. P., Fuesting, M. A., Weisgram, E. S., Claypool, H. M., and Diekman, A. B. (2020), "Putting Belonging in Context: Communal Affordances Signal Belonging in STEM," *Personality and Social Psychology Bulletin*, 46, 1186–1204. [529]

Boucher, K. L., Fuesting, M. A., Diekman, A. B., and Murphy, M. C. (2017), "Can I Work with and Help Others in this Field? How Communal Goals Influence Interest and Participation in STEM Fields," *Frontiers in Psychology*, 8, 901. [529]

Dailey, D., Stone, S., Tanweer, A., and the Data for Good Organizer Network. (2021), "The Data for Good Growth Map: Decision Points for Designing a University-based Data for Good Program," available at *https://escience.washington.edu/wp-content/uploads/2021/11/Data4Good_GrowthMap.pdf*. [530,533]

Davenport, T., and Patil, D. (2012), "Data Scientist: The Sexiest Job of the 21st Century," *Harvard Business Review*. [529]

Dolan, E., and Johnson, D. (2009), "Toward a Holistic View of Undergraduate Research Experiences: An Exploratory Study of Impact on Graduate/Postdoctoral Mentors," *Journal of Science Education and Technology*, 18, 487–500. [529]

Dweck, C. S. (2006), *Mindset: The New Psychology of Success*, New York: Random House. [533]

Fink, L. D. (2013), *Creating Significant Learning Experiences* (2nd ed.), San Francisco, CA: Wiley. [534]

Foster, J. L., Glass, D., Salazar, C., and Tamene, M. (2021), "Geography, Equity, and the Seattle $15 Minimum Wage Ordinance," available at *https://uwescience.github.io/MinWA/*. [534]

GAISE College Report ASA Revision Committee. (2016), "Guidelines for Assessment and Instruction in Statistics Education College Report," available at *http://www.amstat.org/education/gaise*. [529]

Galbraith, M. W. (2003), "Celebrating Mentoring," *Adult Learning*, 14, 2–3. [533]

Garcia, P., Perez, M., Farrell, D., Bork, S., Ericson, B., and Mondisa, J.-L. (2021), "Supporting Mutually Beneficial Near-Peer Mentoring Relationships Within Computing Education Programs," in *2021 Conference on Research in Equitable and Sustained Participation in Engineering, Computing, and Technology*, pp. 1–9. [529]

Haeger, H., and Fresquez, C. (2016), "Mentoring for Inclusion: The Impact of Mentoring on Undergraduate Researchers in the Sciences," *CBE–Life Sciences Education*, 15, 1–9. [533]

Johnson, W. B. (2007), *On Being a Mentor: A Guide for Higher Education Faculty*, New York: Lawrence Erlbaum Associates Publishers. [533]

Lang, J. M. (2016), *Small Teaching*, San Francisco, CA: Jossey-Bass. [533]

Lazar, N. A., Reeves, J., and Franklin, C. (2011), "A Capstone Course for Undergraduate Statistics Majors," *The American Statistician*, 65, 183–189. [530]

Lee, A., Dennis, C., and Campbell, P. (2007), "Nature's Guide for Mentors," *Nature*, 447, 791–797. [533]

Legler, J., Roback, P., Ziegler-Graham, K., Scott, J., Lane-Getaz, S., and Richey, M. (2010), "A Model for an Interdisciplinary Undergraduate Research Program," *The American Statistician*, 64, 59–69. [530]

Leslie, S.-J., Cimpian, A., Meyer, M., and Freeland, E. (2015), "Expectations of Brilliance Underlie Gender Distributions Across Academic Disciplines," *Science*, 347, 262–265. [532]

McGee, M., Stokes, L., and Nadolsky, P. (2016), "Just-in-Time Teaching in Statistics Classrooms," *Journal of Statistics Education*, 24, 16–26. [531]

National Academy of Sciences, Engineering, and Medicine. (2011), *Expanding Underrepresented Minority Participation: America's Science and Technology Talent at the Crossroads*, Washington, DC: The National Academies Press. [529]

——— (2018), *Data Science for Undergraduates: Opportunities and Options*, Washington, DC: The National Academies Press. [530]

National Center for Science and Engineering Statistics. (2022), "Doctorate Recipients from U.S. Universities: 2022," available at *https://ncses.nsf.gov/pubs/nsf24300* [Accessed: September 2024]. [529]

National Research Council. (2001), *Knowing What Students Know: The Science and Design of Educational Assessment*, Washington, DC: The National Academies Press. [534]

Nolan, J. R., McConville, K. S., Addona, V., Tintle, N. L., and Pearl, D. K. (2020), "Mentoring Undergraduate Research in Statistics: Reaping the Benefits and Overcoming the Barriers," *Journal of Statistics Education*, 28, 140–153. [529]

Packard, B. W.-L. (2016), *Successful STEM Mentoring Initiatives for Underrepresented Students : A Research-based Guide for Faculty and Administrators* (1st ed.), Sterling, VA: Stylus Publishing. [533]

Pasick, R., Oliva, G., Goldstein, E., and Nguyen, T. (2010), *Community-Engaged Research with Community-based Organizations: A Resource Manual for Researchers*, Community Engagement Program, Clinical and Translational Science Institute at the University of California, San Francisco. [534]

Pfund, C., Brace, C., Branchaw, J., Handelsman, J., Masters, K., and Nanney, L. (2014), "Mentor Training for Biomedical Researchers," available at *https://www.cimerprojectportal.org/Content/PDFs/completeCurricula/Overview_Biomedical%20Research.pdf*,. [533]

Price, A. M., Kim, C. J., Burkholder, E. W., Fritz, A. V., and Wieman, C. E. (2021), "A Detailed Characterization of the Expert Problem-Solving Process in Science and Engineering: Guidance for Teaching and Assessment," *CBE–Life Sciences Education*, 20, 1–15. [534]

Rodolfa, K. T., Unanue, A. D., Gee, M., and Ghani, R. (2019), "An Experience-Centered Approach to Training Effective Data Scientists," *Big Data*, 7, 246–261. [529]

Russell, L. (2007), *10 Steps to Successful Project Management*, Association for Talent Development. [534]

Schwartz, D. L., Tsang, J. M., and Blair, K. P. (2016), *The ABCs of How We Learn: 26 Scientifically Proven Approaches, How They Work, and When to Use Them*, New York: W W Norton & Co. [534]

Simkins, S., and Maier, M., eds. (2010), *Just-in-Time Teaching: Across the Disciplines, Across the Academy*, Sterling, VA: Stylus Pub. [531]

Slepian, M. L., and Jacoby-Senghor, D. S. (2021), "Identity Threats in Everyday Life: Distinguishing Belonging from Inclusion," *Journal of Personality and Social Psychology*, 12, 392–406. [533]

Smucker, B. J., and Bailer, A. J. (2015), "Beyond Normal: Preparing Undergraduates for the Work Force in a Statistical Consulting Capstone," *The American Statistician*, 69, 300–306. [530]

Sorte, C. J., Aguilar-Roca, N. M., Henry, A. K., and Pratt, J. D. (2020), "A Hierarchical Mentoring Program Increases Confidence and Effectiveness in Data Analysis and Interpretation for Undergraduate Biology Students," *CBE Life Sciences Education*, 19, 1–12. [529]

Stamp, N., Tan-Wilson, A., and Silva, A. (2015), "Preparing Graduate Students and Undergraduates for Interdisciplinary Research," *BioScience*, 65, 431–439. [533]

Steele, C. M., and Aronson, J. (1995), "Attitudes and Social Cognition Stereotype Threat and the Intellectual Test Performance of African Americans," *Journal of Personality and Social Psychology*, 69, 797–811. [533]

Stephens, N. M., Fryberg, S. A., Markus, H. R., Johnson, C. S., and Covarrubias, R. (2012), "Unseen Disadvantage: How American Universities' Focus on Independence Undermines the Academic Performance of First-Generation College Students," *Journal of Personality and Social Psychology*, 102, 1178–1197. [529,532]

Swann, S. A., and Murray, A. R. C. V. J. N. M. C. (2020), "Meaningful Community Collaboration in Research," *BC Medical Journal*, 62, 340–341. [534]

Tenenbaum, L. S., Anderson, M. K., Jett, M., and Yourick, D. L. (2014), "An Innovative Near-Peer Mentoring Model for Undergraduate and Secondary Students: STEM Focus," *Innovative Higher Education*, 39, 375–385. [529]

Trujillo, G., Aguinaldo, P. G., Anderson, C., Bustamante, J., Gelsinger, D. R., Pastor, M. J., Wright, J., Márquez-Magaña, L., and Riggs, B. (2015), "Near-Peer STEM Mentoring Offers Unexpected Benefits for Mentors from Traditionally Underrepresented Backgrounds," *Perspectives on Undergraduate Research and Mentoring*, 4, 1–11. [529]

U.S. Bureau of Labor Statistics. (2025), "Occupational Outlook Handbook: Data Scientists," available at *https://www.bls.gov/ooh/math/data-scientists.htm*[Accessed: May 2025]. [529]

Walton, G. M., and Wilson, T. D. (2018), "Wise Interventions: Psychological Remedies for Social and Personal Problems," *Psychological Review*, 125, 617–655. [533]

Wilson, A. T., and Grigorian, S. (2019), "The Near-Peer Mathematical Mentoring Cycle: Studying the Impact of Outreach on High School Students' Attitudes Toward Mathematics," *International Journal of Mathematical Education in Science and Technology*, 50, 46–64. [529]

Wilson, Z. S., Holmes, L., deGravelles, K., Sylvain, M. R., Batiste, L., Johnson, M., McGuire, S. Y., Pang, S. S., and Warner, I. M. (2012), "Hierarchical Mentoring: A Transformative Strategy for Improving Diversity and Retention in Undergraduate STEM Disciplines," *Journal of Science Education and Technology*, 21, 148–156. [529]

Zaniewski, A. M., and Reinholz, D. (2016), "Increasing STEM Success: A Near-Peer Mentoring Program in the Physical Sciences," *International Journal of STEM Education*, 3, 14. [529]