Inspiring Data Science Interest in Community College Students

Dr. Risa Bobroff Myers, Rice University

Risa Myers is an Associate Teaching Professor in the Department of Computer Science at Rice University, where she teaches students about computer science, databases and data science. She holds degrees from MIT, Stanford, and Rice University and has 20+ years of experience in healthcare, including time at Hewlett-Packard Medical Products Division, Baylor College of Medicine, M.D. Anderson Cancer Center, and Houston Methodist.

Her interests include applications of data science in healthcare and exploring and innovating in data science pedagogy. She does this both through the development of interactive learning materials and by bringing lessons learned from research and her industry experience into the classroom in the form of examples, assignments, and exercises.

Dr. Canek Moises Luna Phillips, Houston Independent School District

Dr. Canek Phillips is a math teacher. His previous experience as an educational researcher included a position as a Research Scientist at in the George R. Brown School of Engineering at Rice University. His research interests are studying the relationship between the ideal of upward mobility and the implications of race, class, and gender in education.

Eric Wuesthoff, Rice University

Eric Wuesthoff is an ecologist and conservation biologist with research and teaching focused in interspecific interactions and ecological restoration. He is currently a doctoral fellow in the Department of BioSciences at Rice University.

Dr. Caroline Masiello, Rice University

Masiello is the W. Maurice Ewing Chair of Biogeochemistry at Rice University, where she directs the Rice Sustainability Institute. From 2018 - 2024 she directed a National Science Foundation Research Experiences for Undergraduates program.

Inspiring Data Science Interest in Community College Students

Introduction

Demand for data science practitioners is growing rapidly, with job growth for individuals with a bachelor's degree in a data science-related field expected to grow at a rate of 35% by 2032 [1]. Authors in journals as diverse as The Harvard Business Review and the IEEE have explored data science career growth and recently upgraded their predictions to an even wider set of industries, with more automated tools and models, and leveraging “low-code/no-code” approaches [2], [3], [4], [5], [6], [7].

There is a particular need for training community college (CC) students in data science as well as in introductory coding and environments, as these students are greater than 40% of the total undergraduate population [8]. Data science and computer science are often taught less rigorously than at 4-year schools, highlighting a need for enrichment in these areas [7]. Furthermore, evidence suggests that for nontraditional students, such as those attending community college, understanding the availability of job opportunities has an outsized effect on their persistence in a major [9], suggesting that their choice to enter and remain in data science would increase if they understood the rapid growth in the field.

Programs that can build knowledge and skills in computing and data science, such as the National Science Foundation’s Research Experience for Undergraduates (REU), are impactful [10], but both time- and cost-intensive for participants and organizers alike. NSF REU programs provide 10 weeks of research experience to ~10 undergraduates per site, typically within a research laboratory at an R1 university, with the goal of increasing students’ likelihood of pursuing STEM careers [11]. However, based on a search of NSF awards with “REU” in the title between January 1, 2020 and December 18, 2025, assuming 10 students per REU per year, the per student cost of running REUs is high (>$12k/student) [12], and program scaling is inherently limited by the number of faculty willing to provide 10 weeks of research mentorship. In addition, the structure of NSF REU programs (duration, hours/day, residential requirements) creates significant logistical challenges for community college students who may simultaneously need to hold jobs and/or care for family members. Finally, in highly diverse cities (e.g., Houston, where no ethnic group is a majority) it is a poor assumption that it is culturally acceptable for these students to move hundreds of miles from home for the summer.

In contrast, we report here on a program piloting a streamlined alternative approach accessible to community college students, a demographic challenging to recruit to NSF REUs.

This experience report describes a pilot program that educates area community college students in data science through the lens of sustainability, using lightweight coding activities accessible to students from any background. The program goals were to broaden students' horizons with regard to data science and sustainability, to measure changes in student scientific identity over time and across a series of workshops, and to strengthen relationships between Rice University and the area community colleges. The goals were to both change student trajectories to be more computational, scientific, or data driven and to increase the number and quality of vertical transfer applicants from the community colleges to four-year universities.

Evidence suggests that providing STEM training in a socially meaningful context enhances persistence. Sustainability has specifically been identified as a learning context that generally helps retention [13]. In the Houston area in particular, sustainability in the context of climate, energy, and environmental health is of great interest: the most recent Rice University Kinder Area Survey found that 84% of residents believed that it is “somewhat likely, likely, or very likely climate change will have a negative impact on their health and well-being in the next 10 years.” This survey has tracked area residents' values since 1982 and reaches more than 5,000 residents each year with an in-depth survey about local values. Since 1982, the survey has found consistent increases in residents' concerns about climate change and its association with human activities [14]. This is not surprising, because climate change impacts residents daily, making roads undrivable due to flooding and making it dangerous to work outdoors in the heat. Therefore, we posited that the sustainability context would be a programmatic pull for area community college students.

Houston, Texas is the 4th largest city in the United States and is considered one of the most diverse cities in the world in terms of racial and ethnic diversity, educational attainment, socioeconomic levels, and culture [15]. As a result of the high immigrant density and broad socioeconomic diversity of Houston, Texas, community colleges play a large role in educating residents. In addition, in 2022, Rice University created a pathway program to help high-achieving community college students successfully vertically transfer to Rice and complete STEM degrees, addressing challenges and needs described by Servin et al. [7]. We leveraged this pathway program to rapidly propose and run our pilot enrichment and outreach program.

Our program's goals were to (1) increase and diversify the number of students with knowledge of and experience in data science and basic coding skills and enable them to identify, understand, and eventually solve, the types of sustainability issues that they see in their communities and in the world at large, to (2) measure any changes in in student scientific identity and to (3) pilot partnerships with area community colleges, all of which are Hispanic Serving Institutions, to build working relationships and better understand student populations.

Related Work

While community college students often choose a major based on highly visible careers (e.g., health professions), they are also motivated by causes close to home, such as a relative suffering from a particular disease or a locally important issue. In particular, Holland Zahner found that students were often motivated by public health and environmental issues [16]. Furthermore, experiences during their education significantly impact students' pursuit of particular majors and careers. For example, showing these students what they can do with a degree in math can open opportunities for them [16]. This characteristic holds true for computing as well. Rodriguez and Ross found that a free course on artificial intelligence at a community college increased student interest in furthering their computing coursework and possible career options [17]. Saidi et al. also highlight the needs and challenges of providing data science educational opportunities at the community college level and specifically highlight the challenge of limited current offerings. These authors also mention how engagement with data science in a non-classroom format can motivate interest in this area [18].

Environmental health, environmental justice, climate change, and the energy transition are all areas of strong interest in this community, and we include them under the umbrella of "sustainability." Data connect each of these topics. For example, literature shows that engagement specifically with environmental health problems increases students' self-efficacy and science identity [19], both of which are correlated with persistence in STEM [20].

The energy transition is a topic of great interest locally and brings a sense of regional opportunity to the discussion of sustainability. Even though the oil and gas industry is a major driver of the regional economy, residents are eager to embrace the energy transition, with more than 70% of residents of all ages agreeing that "Houston, Texas should be a leader when it comes to America's alternative energy future," and "local leaders should incentivize businesses that are focused on producing alternative energy to come to Houston." Among residents aged 18-29 (a major part of the community college demographic) the percentage in agreement with both these statements was 89% [14].

Education literature also points to the effectiveness of the environmental context in teaching STEM concepts. For example, adding environmental context to a general chemistry class increased student motivation and self-efficacy [13], an important predictor of retention in STEM [20]. Other studies have shown improved student performance in introductory classes when context is added [21], with sustainability topics changing students' career aspirations [22]. In environmental fields, outdoor fieldwork is an effective mode of experiential learning that increases student engagement, builds professional competency, and increases retention [23]. However, these longer-duration field experiences are logistically challenging, resource-intensive, and uniquely inaccessible to some students, including many community college students [23], [24]. Data science approaches offer an alternative experiential learning model to educate and train students in sustainability topics and bridge classroom learning to real-world practice.

Here we report the results of a pilot program designed to test the effectiveness of the sustainability context in motivating students to continue in STEM educational tracks. We specifically measured the effects of our intervention against the five metrics associated with persistence in STEM: science identity, scientific community values, self-efficacy, and networking [20].

Program Description

We designed this program with the needs of area community college students in mind. In particular, the program occurred within the spring semester on Fridays, a class-free day for the community colleges. To reduce student commute burden, the instructors traveled to the community college main campuses (typically 25+ miles away), where participants from each colleges' different campuses converged. We conducted the same workshops at each community college systems' main campus.

This program was designed based on two of the authors' experiences attempting to include community college students in traditional, R1-hosted National Science Foundation REU-style programs in two disciplines (computer science and geosciences). While we could recruit hundreds of community college students to NSF REU information sessions, the actual application yield was strikingly low (e.g., 2-3 students/100 recruiting session attendees, even when students were provided with in-session help with program applications). After focused conversations with potential applicants and community college faculty, it became clear that aspects of traditional REU program design made the programs unworkable for many community college students. In particular, the length of programs (10 weeks) and the specific summer dates chosen overlap with portions of two community college summer sessions in our city. In addition, the requirement for a full eight hours/day of activities made it necessary for students to stop any outside work or caregiving, which, despite available stipends, jeopardize students' long-term ability to pay for life necessities. Furthermore, community college student skill sets (e.g., programming experience) were more variable than those of their 4-year college peers, reducing the time or complexity of projects that could be researched. This skill gap would require the REU mentor and student to spend valuable time on skill building or less complex research projects, hindering student integration into the cohort.

We designed our program content depth and duration to accommodate learners earlier in their education. This program was initially conceived as a multi-tiered response to a federal funding call to create programs to train the next generation of data scientists. The proposal had a staged approach comprised of a large-scale introductory session, a workshop series for a slightly smaller number of students, a summer bootcamp, and a research semester, where a reduced number of 'graduates' from each stage progressed through later stages, allowing us to train more students with a broader skill set. At each step, we would assess students' scientific identity and skills. While the program was under review, we had the opportunity to share our ideas with corporate sponsors. While we did not receive federal funding, Google generously supported a pilot.

The topics of data science and sustainability were key to the program, as these are areas of strength for Rice University, and areas where there are often gaps in learning opportunities at the community college level. Furthermore, sustainability was decided upon as the domain because issues such as sustainability are frequently compelling to community college students.
We used the funds from Google to test the impacts of our design's initial workshop series on community college students. We received notification of the gift in November 2024, obtained our Institutional Review Board approval in early January 2025, and held the first workshop four weeks later. The program wrapped up at the end of April, with a poster session for all participants, held at Rice.

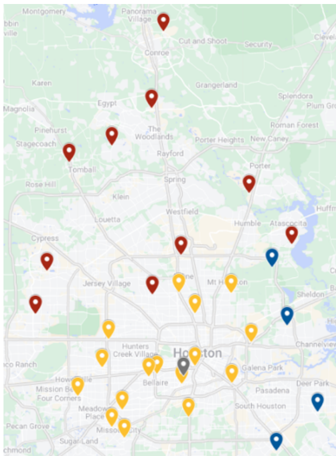
This program and the related surveys were designed to answer the following research questions, which relate to the program goals of increasing diversity and participation in (data) science and measuring changes in scientific identity:

RQ1: Can exposure to data science in the context of sustainability promote students' persistence within the scientific community?

RQ2: How did participants demonstrate cognitive engagement with sustainability and data science curriculum from the workshop series?

RQ3: Did students perceive value from the workshops?

Participants



Students attending the workshop series were recruited from the three large, multi-campus, public community college systems located in [City], [State], a major metropolitan area in the southern United States. See Figure 1 for area locations. Students came from a spectrum of disciplines, including government, business, biology, engineering, and computer science, with a few humanities majors. Admissions were handled directly by the community colleges, and we recommended focusing on students in the honors programs, as this was the first time we were running this type of program. Table 1 shows the student demographics shared with us.

Figure 1. Map of area community college campuses and Rice University. San Jacinto College campuses are blue, Houston City College campuses are yellow, and Lone Star College locations are red. Rice University is located at the gray pin.

Table 1. Participant demographics in percentage representation for each school.

Demographic	Details	CC1	CC2	CC3
Female		30%	70%	42%
Male		70%	30%	58%
Race/Ethnicity	Asian	Included in Non-Hispanic	54%	Not available
	Black	Included in Non-Hispanic	25%	Not available
	Hispanic	65%	13%	Not available
	Non-Hispanic	35%	8%	Not available
Majors	STEM	78%	75%	69%
Honors		48%	100%	100%

Based on the available budget, 28 students from each community college were invited to participate. Of those 84 students, 77 attended the first day of the program. All students who attended the second day of the program were given a stipend, a portion of which was delivered mid-program. Students who successfully completed the program were awarded the balance of their stipend as well as a micro-credential in Environmental Data Science from our university's

School of Continuing Studies, with whom we partnered to deliver this program. In total, the overall research cohort ended with 73 participants.

Workshops

The topics for these workshops were chosen both based on relevance to local sustainability challenges and by availability and interest of faculty, postdoctoral scholars, and graduate student instructors, who needed to be recruited and trained within the very short time between funding delivery and program commencement. The program started with a kick-off event at each campus, consisting of an overview of data science and sustainability, an introduction to using data to help communities address a regional groundwater contamination-driven cancer cluster, and a one-hour workshop on the environmental impact of the Internet (e.g., data centers and subsea cables). These topics and speakers were intentionally selected to start the program with experienced instructors as well as to serve as an introduction to environmental topics related to technology, setting the stage for future workshops.

Following the kick-off, there were two, 2 hour Friday workshops focused on data science skills. We addressed food insecurity by providing time-of-day appropriate food. During each workshop the instructor began with an introductory lecture, and then students completed activities that required them to review data and run Python code, including generating visualizations. All the code was run using Jupyter Notebooks on Google Colab, requiring students to have an internet accessible device. All students were able to bring and use either laptops or tablet computers. We distributed files to students initially through Box folders and later through our university's learning management system.

The second workshop instructor was selected due to his dynamic mentoring and lecturing style. Professor César Uribe introduced his work on food webs and used this engaging topic to guide students through a fill-in-the-blank Jupyter Notebook, where they explored different data representations and visualizations to help determine the consequences of applying limited conservation funding to preserve or protect one species in a food web over another. The third workshop focused on data visualization, using local air quality data. Again, students were introduced to the material and worked with fill-in-the-blank notebooks using supplied data files. The data visualization instructor was recruited because she is an experienced and popular lecturer, and both the topic of local air quality and data visualization were impactful and interesting.

The final workshop of the program focused on applying data science concepts and practice to a diverse set of topics in sustainability. The final workshop, which, like the previous program session was repeated three times for each community college cohort, was taught by three instructors over a period of six hours and offered catered breaks and meals. We recruited and trained graduate student and postdoctoral scholar instructors from Rice University for this final workshop. All trainee instructors received eight hours of training from the program PIs and from the Rice's teaching and learning center on developing clear learning outcomes and presentation skills.

Each of three modules during this final workshop was taught through interactive, active-learning lessons using a mix of informational lectures, data analysis demonstrations, and participatory exercises using provided data from instructors and community science sources. Each module was taught by a lecturer directly conducting research on the module topic and using data science in their scholarship. These sustainability workshops explored the role of microbial wastewater sampling to inform environmental public health, techniques to model hazards to achieve sustainable space travel, and applications of community science data to explore biodiversity in rainforests.

The first module centered around microbial pathogen detection in municipal wastewater samples, with key relevance to local environmental and public health challenges including recent regional outbreaks of coronaviruses and measles. Students were taught about major variables that impact viral detection and practiced assessing presence of agents and calculating concentrations. The following module in the final workshop discussed concepts of sustainability in space travel and covered assessing and mitigating hazards to astronauts during missions. With instructor guidance, students built modules of increasing complexity to evaluate oxygen conservation in spacecraft and calculate thresholds for safe missions in the event of oxygen loss in emergencies. Finally, in the last module students learned about the application of community science to biodiversity monitoring and the interaction of species and resources in complex ecosystems. The instructor gave examples from his PhD research on tropical seed dispersal by lemurs and then taught students to model species interactions using publicly available ecological data from the iNaturalist app.

Poster Session

At the end of the program each student created and presented a 24" x 36" reflection poster describing how the program changed their views or career plans. We provided a poster template and recommended that students keep an online journal during the program. To help those unfamiliar with making this type of poster, we held and recorded an online session about what a reflection poster is and how to create one. Each college printed their students' posters.

The participants from all three community colleges attended the same capstone poster session, which was organized as a celebratory event. There were snacks and poster boards set up for displaying multiple posters, and we assigned students from different schools to the same board to encourage inter-school interaction. We also staggered the time when students were expected to be at their posters to allow them time to circulate and make professional connections.

Program Assessment Methods

We administered a web-based survey at the end of each workshop, consisting of quantitative Likert-style responses, giving students the option to opt out of participation in any survey. We used four constructs in building the quantitative part of the survey. The bulk of the survey contained items from the Persistence in the Sciences (PITS) instrument [20], a validated tool (Cronbach's alpha = 0.96) widely used in STEM education research to assess psychosocial factors linked to persistence in science. The PITS survey was originally designed within the context of a research experience for 4-year college students, which made some items inappropriate for our more limited program. However, we elected to use it, as one of the ancillary goals of our program was to assess which of the gains from longer, more traditional outreach programs could be met with shorter-term programs designed specifically for community college students.

We selected 15 total items from four of the six PITS survey constructs: Science Identity, which measures the level to which the participant would call themselves a scientist (five items); Networking, which measures participant ratings pertaining to social aspects of major scientific activities (three items); Science Community Values, which measures participants' sense of affinity towards scientific cultural practices (three items); and Self-Efficacy, which measures participants' sense of capability towards performing scientific tasks (four items). We added a Workshop Satisfaction construct to measure the extent to which participants would say they were satisfied with different aspects of the workshop (11 items).

We also included three qualitative short answer items at the end of each survey based on the theme of each workshop (e.g., Society and Sustainability, Food Webs, Data Visualization). Example survey questions are posted in Table 2. We coded the free responses to these questions based on Bloom's Taxonomy of Knowledge to assess participants' learning [25]. Bloom's taxonomy describes levels of knowledge, ranging from concrete forms, like recalling information, to higher level abstract thinking, e.g., synthesizing knowledge.

We used SPSS for quantitative data analysis, including Cronbach's reliability scales and Friedman's tests for possible differences across workshops. Friedman's tests are dependent measures, meaning each participant's responses are compared to their own responses over time. Thus, any significant differences reflect whether participants changed their own rankings across surveys. We also used a post hoc Dunn's test to determine where those changes occurred.

We tested all quantitative items with non-parametric tests because they do not require data that fit normality assumptions [26]. Each construct was tested for internal consistency using Cronbach's alpha, which measures the reliability of the survey instrument and the independence of the survey questions. Items with no response or marked as "Not Applicable" were excluded from the analysis.

For Friedman's tests, exclusion diminished overall sample sizes because they are pairwise tests.

Table 2. Example Items from Each Survey Construct [20]

Construct	**Example Item**
PITS Science Identity	I think of myself as a "scientist."
PITS Networking	I have discussed my participation in this program with people in my community.
PITS Science Community Values	I am a person who thinks that scientific research can solve many of today's world challenges.
PITS Self-Efficacy	I am confident that I can generate a research question that data science can answer.
Sustainability Impact	The sustainability focus of the Sustainability Data Science School was a major reason for my enrollment in the program.

Workshop Evaluation	My knowledge of data science increased as a result of this workshop.
Learning Objective	Identify at least two factors that could lead to an increase in the detected viral concentration in wastewater.

Analysis and Interpretation

For two of the constructs (Table 3) – Science Identity and Workshop Evaluation – our tests detected significant gains between the first and last surveys. Averages improved from 4.27 to 4.49 and from 4.52 to 4.74, respectively.

Table 3. Non-Parametric Tests. Participant Changes (n - Number of participants; chi squared – Overall significance in difference between any survey; degrees of freedom – The default amount of independent information in the dataset was always 3) Overall (Z – Specific significance level between first and last survey; and Standard Error (SE))

	Participant Changes: Friedman's Test				**Overall Comparison: Post-hoc Dunn's Test**			
Construct	**n**	**Chi sq**	**p-val**	**Effect Size (W)**	**Z**	**SE**	**p-val**	**Adj. p-val**
Science Identity	49	11.45	< 0.01	0.0780	-0.735	0.261	< 0.01	< 0.05
Networking	49	8.46	< 0.05	0.0576	-0.582	0.261	< 0.05	0.1540
Science Community Values	47	1.55	0.6708	0.0110	-0.234	0.266	0.3795	1.0000
Self-Efficacy	50	13.10	< 0.05	0.0873	-0.650	0.258	< 0.05	0.0709
Workshop Evaluation	45	22.09	< 0.001	0.1637	-1.067	0.272	< 0.001	< 0.001

For these two constructs, we observed a slight effect size, which we interpret as a promising outcome given that many participants began with high baseline scores. This limited the room for measurable growth but still suggests meaningful gains in areas central to the program's goals.

Friedman's tests are used to identify areas where significant changes occurred in the surveys. We followed up with Dunn's pairwise comparisons to identify where the significant changes were actually present: In both scientific identity and workshop evaluation.

A further test, Kendall's (W) Effect Size, describes the magnitude of the changes. Per standard ranges, our results indicate modest changes (see Table 3). For our two significant tests, our highest values were 0.0780 and 0.1637, which are considered weak changes in the magnitude of responses, indicating a modest change [27].

Survey Results

Science Identity

Scientific identity is crucial to persistence within the scientific community. During our program, students' perception of themselves as (data) scientists increased, as seen by a significant change in the Science Identity construct from the survey after the first workshop to the survey after the final workshop.

Only one item in the construct of Science Identity showed an increase over the course of the program. "I think of myself as a 'scientist'" had the lowest initial values in the first survey and displayed a statistically significant mean increase by the last survey, increasing from 3.65 to 4.39 (out of 5). Other items within this construct were not significant on their own.

We believe we had more room for growth with this item because of the societal prestige and sense of meritocracy surrounding the idea of being a scientist. It is possible that students did not feel like they could call themselves a scientist without working in a community where they could make conclusions about their data and have the results respected. We hypothesize that the participation in a program run by a high-status university both increased their knowledge and gave them confidence to identify as a member of the scientific community.

Per the Dunn's test, our data show that the improvement in this item's ranking was positive but gradual over the course of the workshops. This demonstrates the effectiveness of our approach in promoting participants, even with high persistence markers in other areas, in building their sense of scientific identity. We hypothesize that the effect size would be larger if we are able to grow our participant pool to a larger set of students not predominantly in the Honors Colleges.

It is particularly notable that the persistence construct of 'scientific identity' grew in our students. In other studies, this is one of two parameters that have played the most significant role in students' intention to stay in STEM [28].

Institutional differences

Within workshops, we also performed statistical tests (Kruskal-Wallis (KW)) to determine if there were significant differences between participants by institutional affiliation for the PITS and Workshop evaluation constructs. We performed tests with post-hoc pairwise comparisons, a standard approach to determine differences between groups using non-parametric data. Like Friedman's tests, the KW test uses a test statistic to calculate an associated p-value for significant difference between groups. Kruskal-Wallis (KW) post-hoc pairwise comparisons show that in Workshops I and II participants' responses for the Science Communication item: 'I have discussed my participation in this program with people in my community', was significantly lower for participants in CC1 compared to participants in CC2. (Workshop I: $n=73$, adj. p-value < 0.05, epsilon squared = 0.366); Workshop II: $n=69$, adj. p-value <0.05, epsilon squared = 0.394). There was no significant difference between CC1 and CC3. These differences may be attributable to different interpretations of the survey questions, which could be related to demographic and geographical differences between schools. In particular, CC2 serves an area with higher population density and a larger international population, while CC1 and CC3 are more suburban in nature.

Subject matter engagement

Participants responded to three open-ended items at the end of each workshop survey—twelve in total —with prompts such as “What are the key social impacts of data centers?” and “How can we leverage data science to enable conservation in ecology?" Two researchers evaluated student responses for demonstrated engagement (0 = No Engagement, 1 = At least Some Engagement) and cognitive level based on Bloom’s Taxonomy. Over 90% of responses across items showed some evidence of understanding, except for the data centers question, which received lower ratings (77% and 88%) and more vague answers.

We calculated interclass correlation coefficients to measure the quality of interrater reliability, which ranged from 0.604 to 1.000. One item showed excellent inter-rater reliability (1.000), and the rest fell within the moderate to good range (0.609 to 0.886), indicating generally consistent scoring between reviewers [29]. Most student responses were classified at the Remembering (43%–96%) or Understanding (4%–57%) levels of Bloom’s Taxonomy, with only 0%–6% reaching the Applying level.

Taken together, the end of survey free responses demonstrated that it was a good instrument to elicit lower-order thinking. This finding may reflect survey fatigue, available time, question wording, program delivery, etc. In future iterations, the team should include activities and ask questions using keywords that encourage higher level thinking.

Workshop evaluation

We measured whether the students found the workshops to be valuable, and for all workshops, all item ratings were consistently at mean levels above 4. We observed our highest effect size and adjusted significance level for the workshop evaluation construct compared to all other constructs ($W=0.1680$, adj. p-val<0.001).

Our evaluation indicated that we were able to improve the experience of participants as the workshops proceeded. Further post-hoc analysis showed that our greatest improvement was between Workshops I and II. The initial workshop included administrative tasks (e.g., paperwork to award stipends). It is likely that the improved experience was a result of the absence of paperwork and enhanced content in the workshops that followed.

Summary

The significant overall changes in participant responses reflect the program’s success in achieving our intended outcomes of increasing markers of persistence and satisfaction in program participants. The results are especially notable because most participants were honors college students, whose markers for persistence in the sciences may have already been higher than the typical student at their institutions. Indeed, in our initial workshop survey, students already marked many of the markers of persistence at a rank of somewhat agree (4) or strongly agree (5), making it harder to achieve improvement of persistence rankings [30]. That we still achieved significant overall improvement in one of the four persistence constructs, and the evaluation suggests that the intervention was impactful even among students who already demonstrated strong baseline indicators of persistence—underscoring its potential effectiveness across a broad range of student readiness levels [30].

Internal Reliability of Constructs

To ensure the validity of our findings, we first calculated Cronbach's non-standardized or "raw" alpha for each survey construct, then applied the non-parametric tests to evaluate changes to students' belonging and workshop effectiveness across the workshop series and to explore differences across institutions within workshops.

Cronbach's alpha is a scale between 0 and 1, where values above 0.7 indicate that survey questions are reliable [30]. All items were on 1-5 Likert scale, so using the non-standardized value is sufficient in the case of our survey. All constructs exceeded this conventional threshold, as shown in Table 4. The Science Identity construct is above 0.9, indicating that we should subtract items in future surveys [30]. The high internal consistency and reliability of the survey constructs bolster the credibility of our analysis.

Table 4. Construct Mean Value and Reliability (Cronbach's alpha)

Construct	Mean	Raw Alpha
PITS: Science Identity	4.36	0.938
PITS: Science Community Values	4.69	0.839
PITS: Networking	4.17	0.844
PITS: Self-Efficacy	4.32	0.881
Workshop Evaluation	4.64	0.895

In Table 4, we also report the mean scores for each construct, aggregated across all workshops. Means for the items of each construct indicate that responses were generally between Somewhat Agree and Strongly Agree.

Participants

Table 5 shows the percentage of admitted students from each community college who attended the program's first day and completed it (in part or in full).

Table 5. Program completion rates by community college. Note that numbers may not sum correctly due to rounding.

	CC1	CC2	CC3
% of admitted students who attended Workshop 1	82	93	100
% of admitted students who partially completed the program	4	21	0
% of admitted students who fully completed the program	79	64	93
Total % of admitted students who completed the program	82	86	93

Reflection and Next Steps

Our analyses of increases in students' persistence markers indicate program success despite being planned and executed very rapidly. A key factor was early engagement by institutional leadership (Vice Chancellors, Honors College directors) at all the area community colleges. These administrators had the authority to commit to participation with little to no additional approvals required. In addition, each school provided dedicated staff who ensured that key milestones were met. In-house partners, such as the Rice University's Glasscock School of Continuing Studies provided learning management system access for the program participants, were also important.

Programmatic flexibility was also an important factor in this program, allowing us to integrate program changes in response to community college partner input and needs. In particular, one of the community college leaders suggested adding the reflection poster component to the program. The reflection posters required the students to consider the impact of the program on their futures and share those thoughts in a tangible way. The poster session also gave the students the opportunity to meet other participants and visit the Rice University campus.

Taking the program to the community colleges was crucial for this demographic because it allowed us to provide an accessible, in-person program. Two of the designers of this program are experienced with online education and recognize that an online program might be more convenient. However, the in-person element was essential for building community and facilitated real-time checking student understanding. Equally important was bringing the students to Rice at the end of the program. For some participants, this was the first time they had set foot on a 4-year college campus.

Shared meals occurring during lunch and snack breaks were also good opportunities for the participants to engage more informally with the instructors. Food insecurity is common among community college students and studies show that people learn better when they are not hungry [32], [33], [34], [35]. In-person programming allowed us to ensure that students' nutritional needs were met during class days.

The four workshop dates and poster session spread across a three-month period worked well for the students, as it was easier for everyone to work around other commitments. In addition, containing the program within a single semester simplified logistics. Ideally, recruitment and admissions would be handled the semester prior to the program.

The stipend component of the program was also appreciated by students. Community college students may face different financial hardships compared to traditional students, and may have additional demands on their time, such as caregiving. To reduce barriers to participation, we provided a stipend delivered in two phases: once in the form of a debit card at the start of the program and the other as a bank transfer when the program was successfully completed. The initial debit card amount was $100 and was intended to help cover incidental expenses during the program (e.g., parking at Rice for the poster session). We found that distributing the debit cards at the second workshop was optimal, as a small number of students did not return after the first workshop, allowing us to reallocate funds.

The vast majority of the students completed the program and received their stipend balance. In a few cases, students either did not attend a significant portion of the workshops or did not submit a reflection poster. These students saw a reduction or elimination in the balance of their stipend. The internal mechanics of stipend disbursement were a challenge for our institution, as it was a complex, multi-step process. Going forward, we would transfer funds at the end of the program to foundations associated with each community college, which could then distribute the funds to participants. This approach may also reduce or eliminate overhead costs.

Next Steps

We hope to repeat this program, adding participant prep work prior to each workshop. We also hope to hold a follow-up, intensive coding camp for a subset of participants. This type of program has been successful in knowledge gain in similar communities [36]. We plan to track student success via LinkedIn, status updates from their home institutions, and, funding permitting, via state and national clearinghouse databases.

Ideally, we would have had a knowledge assessment survey prior to the first workshop and after the poster session. This was not possible due to the short timeframe of the program. Similarly, we should include a final student survey to assess scientific identity, knowledge, and satisfaction with the overall program. Future programs will include such components to measure actual learning.

Based on the institutional differences we found in the Science Communication construct, we also plan to find ways to build in specific cues within the curriculum during the first workshop to address the importance of speaking about scientific activities with community members, as well as expanding on the concept of community. Doing so would reflect the importance of the activity in the development of persistence. We should also add the item to cognitive interviews to better understand the way students are interpreting the question and be conscientious of what culturally relevant communication can be in the student context. For instance, how students regard members of their community may greatly differ, and we may need to add language so that students better understand what potential opportunities for communication could manifest so that they are more aware during the program and surveys.

Conclusion

We were able to rapidly design and implement a data science and sustainability STEM persistence program tailored to community college students’ pedagogical needs. Despite the abbreviated timeframe available for program construction, our approach was highly impactful, as evidenced by statistically significant growth in students' scientific identity. We also saw growth in responses to questions about the material presented, and in enthusiastic and qualitative reflections on their end-of-program posters. Approaches such as those used in this program have the potential to strengthen community college students' scientific interests and computational skills at a relatively low cost in terms of time and money compared to REU style programs.

We found that integrating sustainability and data science in a workshop series can create meaningful educational, scientific, and computational experiences for students, particularly along dimensions of identity that foster a sense of belonging in science and promote future interest in data science, computing, and sustainability.

We were also successful in introducing students to basic coding concepts and tools, enabling them to alter and run code, regardless of their backgrounds.

Finally, the response from the students and institutional leadership about the program was both statistically significant and anecdotally overwhelmingly positive. Some student quotes include:

“This experience reshaped how I think about environmental data - not as abstract numbers, but as a foundation for real-world decision-making. I now see data science as a tool I want to keep using in school, in research, and possibly in a career that blends sustainability with data-driven solutions.”

“I now see technical skills not as a means to the end of analysis, but as being at the core of creating actionable, measurable change in the messy systems we are working in.”

We statistically tested a number of aspects of both our PITS and workshop evaluation constructs, and our results indicate that even though we worked with multiple community colleges, we were able to deliver material to similar satisfaction levels at all three institutions. In addition, our analyses appear to have detected modest shifts in student satisfaction associated with administrative burdens, leading us to the unsurprising conclusion that relieving students of administrative tasks improves their satisfaction with academic programming.

Overall, pairing sustainability with data science resulted in a different set of students than we otherwise would have engaged. We had the opportunity to, and did, change lives.

Acknowledgements

We express our gratitude to Google, in particular to Partha Ranganathan, Sepi Hejazi Moghadam, Chloe Tims Wright, and Ranier Aliment, for seeing the spark in this program and making it possible. Our community college partners provide insight, found us fantastic students, and were a joy to work with: Associate Vice Chancellor Katharine Caruso, Elizabeth Alexander, and Jessica Miller at Lone Star College; Associate Vice Chancellor Miguel Ramos and Dr. Darin Baskin at Houston City College; and Associate Vice Chancellor Rachel Garcia and Janna Lasater at San Jacinto College.

We would also like to thank our wonderful and knowledgeable instructors who shared their enthusiasm and expertise with our students: Dr. Julia Huggins, Assistant Professor César Uribe, Assistant Teaching Professor Marmar Orooji, Dr. Jingjing Wu, Bjorn Shockey, and Eric Wuesthoff. Furthermore, we could not have managed the program without the able assistance of Gabriela Mendoza and Maria Martinez. Dr. Mary Glavan provided editing expertise to help us share our story, and Leah Aschmann helped make it possible.

References

[1] U.S. Department of Labor Bureau of Labor Statistics. "Occupational Outlook Handbook, 'Data Scientists'." https://www.bls.gov/ooh/math/data-scientists.htm [accessed 6/12/2025].

[2] N. Ahmad, A. Hamid, and V. Ahmed, "Data science: Hype and reality," *Computer*, vol. 55, no. 2, pp. 95-101, 2022.

[3] N. Ahmad and A. Hamid, "Will Data Science Outrun the Data Scientist?" *Computer*, vol. 56, no. 2, pp. 121-128, 2023.

[4] T. H. Davenport and D. Patil, "Is data scientist still the sexiest job of the 21st century," *Harvard Business Review*, vol. 15, 2022.

[5] J. Malyn-Smith, S. MacGillivray, M. Harris, and J. Polzin, "Developing New Data Pathways in Community Colleges," in *SITE Interactive Conference, 2020*: Association for the Advancement of Computing in Education (AACE), pp. 360-365.

[6] T. Patil and T. Davenport, "Data scientist: The sexiest job of the 21st century," *Harvard Business Review*, vol. 90, no. 10, pp. 70-76, 2012.

[7] C. Servin, E. K. Hawthorne, L. Postner, C. Tang, and C. S. Tucker, "Curricular and pedagogical considerations in computer science education: The role of community colleges for the next decade," in *Proceedings of the 55th ACM Technical Symposium on Computer Science Education* V. 1, 2024, pp. 1196-1201.

[8] Community College Research Center, "An Introduction to Community Colleges and Their Students," Columbia University, 2025. [Online]. Available: https://ccrc.tc.columbia.edu/publications/introduction-community-colleges-students.html [accessed 6/12/2025].

[9] G. R. Stuart, C. Rios-Aguilar, and R. Deil-Amen, "“How much economic value does my credential have?” Reformulating Tinto's model to study students' persistence in community colleges," *Community College Review*, vol. 42, no. 4, pp. 327-341, 2014.

[10] Y. Jin, C. Qian, and S. Ahmed, "Closing the Loop: A 10-year Follow-up Survey for Evaluation of an NSF REU Site," in *2022 ASEE Annual Conference & Exposition*, 2022.

[11] NSF. "NSF REU: Information for Faculty." https://www.nsf.gov/funding/initiatives/reu/faculty [accessed 6/12/2025].

[12] NSF REU awards https://www.nsf.gov/awardsearch/search-results?BooleanElement=All&BooleanRef=All&Keyword=REU&awardTitleOnly=true&expiredAwards=true&resultsPerPage=100&sortBy=date [accessed 6/12/2025].

[13] O. Gulacar, C. Zowada, S. Burke, A. Nabavizadeh, A. Bernardo, and I. Eilks, "Integration of a sustainability-oriented socio-scientific issue into the general chemistry curriculum: Examining the effects on student motivation and self-efficacy," *Sustainable Chemistry and Pharmacy*, vol. 15, p. 100232, 2020.

[14] Potter, D., Pren, K., Tobin, A., Perez, K., Njeh, J., Glanzer, A., Williams, L., Niznik, A., Dawson, L., & Dulin, M. (2024). The 43rd Kinder Houston Area Survey: Houston of Tomorrow: Perspectives from a city shaping America's future.

[15] A. McCann. "Most Diverse Cities in the U.S. (2025)." https://wallethub.com/edu/most-diverse-cities/12690 [accessed May 14, 2025, 2025].

[16] D. G. Holland Zahner, "Navigating STEM major and transfer destination choices: Community college student experiences through the lens of practice theory," *Community College Review*, vol. 51, no. 4, pp. 538-566, 2023.

[17] S. L. Rodriguez and A. Ross, "How offering a free introductory AI computing course at a hispanic-serving community college influences computing interests for enrolled students?" *Computer Science Education*, vol. 35, no. 3, pp. 523-549, 2025.

[18] R. Saidi, H. Burn, B. Baumer, J. Black, D. Deb, and Z. Drozda, "Building Ramps and Pathways to Data Science in K-12, Minority-Serving Institutions, and Two-Year Colleges," *Scatterplot*, vol. 1, no. 1, p. 2395091, 2024.

[19] S. J. Adkins-Jablonsky, R. Akscyn, B. C. Bennett, Q. Roberts, and J. J. Morris, "Is community relevance enough? Civic and science identity impact of microbiology CUREs focused on community environmental justice," *Frontiers in Microbiology*, vol. 11, p. 578520, 2020.

[20] D. I. Hanauer, M. J. Graham, and G. F. Hatfull, "A measure of college student persistence in the sciences (PITS)," *CBE—Life Sciences Education*, vol. 15, no. 4, p. ar54, 2016.

[21] S. I. H. Sristy, C. Z. Muteti, Y. Vasquez, and J. M. Mutambuki, "Integrating Real-World Contexts and Application Cards Activities in the General Chemistry I Course Increases Students' Achievement Scores and Perceptions of Chemistry Relevance," *Journal of Chemical Education*, vol. 100, no. 12, pp. 4608-4618, 2023.

[22] O. Gulacar, R. Marwaha, and K. R. Goradia, "Examining changes in students' perception of science relevancy and their career aspirations: Integrating sustainability-oriented socio-scientific issues into general chemistry curriculum," *Sustainable Chemistry and Pharmacy*, vol. 25, p. 100577, 2022.

[23] B. Munge, G. Thomas, and D. Heck, "Outdoor fieldwork in higher education: Learning from multidisciplinary experience," *Journal of Experiential Education*, vol. 41, no. 1, pp. 39-53, 2018.

[24] A. N. Robin, A. A. Farmer, K. O'Connell, A. K. Varty, J. A. Hewlett, and J. W. Lee, "Community College Students in the Field," *Bulletin of the Ecological Society of America*, vol. 103, no. 3, pp. 1-10, 2022.

[25] B. S. Bloom, M. D. Engelhart, E. J. Furst, W. H. Hill, and D. R. Krathwohl, *Taxonomy of educational objectives: The classification of educational goals. Handbook 1: Cognitive domain*. Longman New York, 1956.

[26] J. de Winter and D. Dodou, "Five-Point Likert Items: t test versus Mann-Whitney-Wilcoxon (Addendum added October 2012)," *Practical Assessment, Research, and Evaluation*, vol. 15, no. 1, p. 11, 2010.

[27] J. R. Landis and G. G. Koch, "The measurement of observer agreement for categorical data.," *Biometrics*, vol. 33, no. 1, pp. 159--174, 1977.

[28] M. Estrada, A. Woodcock, P. R. Hernandez, and P. Schultz, "Toward a model of social influence that explains minority student integration into the scientific community," *Journal of Educational Psychology*, vol. 103, no. 1, p. 206, 2011.

[29] T. K. Koo and M. Y. Li, "A guideline of selecting and reporting intraclass correlation coefficients for reliability research," *Journal of Chiropractic Medicine*, vol. 15, no. 2, pp. 155-163, 2016.

[30] L. C. Johnson, R. Beaton, S. Murphy, and K. Pike, "Sampling bias and other methodological threats to the validity of health survey research," *International Journal of Stress Management*, vol. 7, no. 4, pp. 247-267, 2000.

[31] M. Tavakol and R. Dennick, "Making sense of Cronbach's alpha," *International Journal of Medical Education*, vol. 2, p. 53, 2011.

[32] K. E. Baugus, "Food insecurity, inadequate childcare, & transportation disadvantage: Student retention and persistence of community college students," *Community College Journal of Research and Practice*, vol. 44, no. 8, pp. 608-622, 2020.

[33] R. T. Ilieva, T. Ahmed, and A. Yan, "Hungry minds: Investigating the food insecurity of minority community college students," *Journal of Public Affairs*, vol. 19, no. 3, p. e1891, 2019.

[34] M. E. Maroto, A. Snelling, and H. Linck, "Food insecurity among community college students: Prevalence and association with grade point average," *Community College Journal of Research and Practice*, vol. 39, no. 6, pp. 515-526, 2015.

[35] J. A. Innis, M. Bishop, and S. Boloudakis, "Food insecurity and community college students," *Community College Journal of Research and Practice*, vol. 44, no. 9, pp. 694-699, 2020.

[36] E. Pearson, V. Latson, and P. Morgan, "Promoting Data Science to Underrepresented Communities Through the Use of Active Learning," in *2023 Congress in Computer Science, Computer Engineering, & Applied Computing (CSCE)*, 2023: IEEE, pp. 1068-1074.