Research Statement

Suraj Rampure

A driving principle in my academic career thus far has been to share as much as possible with as many people as possible, so that students and educators across the world – not just at my institution – can benefit from the work I have done. I have made it a point to ensure that all course materials I have created are free for anyone on the internet to access; this includes slides, videos, and assignments in my Introduction to Mathematical Thinking course (link), offering of Principles and Techniques of Data Science (Data 100) in Summer 2020 (link), and soon-to-be-offered Introduction to Computational Thinking with Data course (link). Moving forward, I intend to keep all resources I create public, and am eventually interested in creating online courses for the specific purpose of public consumption.

My first foray into formal academia came in June 2020 when I gave several presentations at the National Workshop on Data Science Education (NWDSE). NWDSE is a UC Berkeley-hosted conference (link) for educators from universities across the world who have implemented or are looking to implement data science curriculum at their institutions. I created a pre-recorded video (link) on the content and structure of Data 100 along with Joseph Gonzalez (the creator of the course). I also co-hosted a Q&A session (link) with Fernando Pérez (another Data 100 instructor and creator of Project Jupyter) where we answered questions about logistical and pedagogical considerations made in the design of Data 100; for instance, "How much calculus should we require students know coming in?" Lastly, I ran a workshop (link) in which I demoed several Data 100 assignments to over 50 participants, and gave them insight on the assignment design and autograding processes for the course.

In my final undergraduate year, I started to engage with the academic community more broadly; I was slated to give one of the aforementioned NWDSE presentations at SIGCSE '20 along with other member's of Berkeley's Data Science Education Program before the symposium was cancelled. Along those lines, however, I recently co-authored a paper that detailed all of the experiments we tried in the Summer 2020 offering of Data 100, the first fully-remote offering of the course. In the paper, we detailed our approaches to the design of lectures, discussion and lab sections, and exams, and describe the perceived effectiveness of these experiments from students in the course. The paper, "Experiences Teaching a Large Upper-Division Data Science Course Remotely", was accepted to SIGCSE '21; a near-finished draft can be provided upon request. As the beneficiary of several other educators' reflections, I think it is imperative that we continue to share what does and doesn't work in our courses with others so that nobody feels the need to re-invent the wheel.

In addition to sharing curriculum and course design openly, I am interested in conducting formal education research. Just before the Summer 2020 semester, I submitted a request to Berkeley's Institutional Review Board to conduct a randomized control trial on students in our class. Specifically, I wanted to determine if students who completed our proof-based linear regression homework on paper would have a different level of understanding of regression than students who completed the same homework problems in the Jupyter notebook environment. The plan was to randomly assign consenting students to one of the two assignment formats and look for trends in their exam scores on relevant problems. While we didn't get a chance to run

the experiment, I remain interested in exploring different assignment formats in data science courses, and would like to run similar experiments in the future. Furthermore, I am interested in analyzing the impact of various initiatives – major GPA caps, new introductory courses, and small group tutoring, for instance – on student achievement and confidence. I have begun to explore data pertaining to my Introduction to Mathematical Thinking course; specifically, I am analyzing anonymized student transcripts to determine whether or not the course made an impact on students' grades in further coursework. I find such analyses to be crucial to the upkeep of a degree program, as they tell us what interventions have a positive impact on students.

As I progress into the next stage of my career, I would like to continue to teach in the public domain, share what I have learned as an instructor with my peers, and conduct more formal computing education research. The collaborative nature of the Halıcıoğlu Data Science Institute at the University of California, San Diego will allow me to pursue these interests alongside other education-focused faculty and students.