The Utility of Designing Data Science Education Programs from a Framework of Identity

June Ahn, Seth Van Doren, Jessica Cai, Ha Nguyen, Fernando Rodriguez,
Christopher Martinez, Jenny Han

Introduction

A fundamental conundrum in data science education is understanding how to foster equitable outcomes in the field. One specific equity concern involves the stark gap between the substantial opportunity that is rising with the growing availability of data and its influential role in all aspects of society, contrasted with the inequitable opportunities for learners to be a part of this growing field. Data has become easier to collect, store, connect, aggregate, analyze, and integrate into a wide variety of professions or occupational pathways (Persaud, 2021). In this new world, data is a vitally important tool and currency. Immense power and opportunity is afforded to those who are literate with data, or even further, those who can work with data in increasingly more advanced and ethical ways (Lang et al., 2017; Siemens & Baker, 2012). The demand for data scientists continues to grow with diverse new career opportunities and roles proliferating rapidly (Davenport & Patil, 2022). This demand is likely to be met in inequitable ways. If history offers any lessons, students from more privileged backgrounds will likely be better positioned to take advantage of these emerging opportunities. How might we explore this conundrum, defined as the gap in opportunity versus equitable inclusion into the field?

In the following chapter, we outline an analytical lens that focuses on exploring how learners can come to identify with the field of data science, and in particular the process through which learners specifically might develop a deeper identity to **become a data scientist** as they progress through their lives. This focus on identity development complements other potential ways to define an equitable outcome. For example, research in data science education often focuses on core knowledge and skills that are part of working with data (Kim, 2016; Persaud, 2021), expanding what we mean as being literate with data to move beyond just formal skills to also value personally relevant ways of engaging with data (Clegg et al., 2020; Clegg et al., 2022), or figuring out ways to better integrate data work with social justice or humanistic experiences to hopefully connect with more learners (Calabrese Barton et al., 2021; Lee, Wilkerson, & Lanouette, 2021). We categorize this focus as attending to instructional concerns or identifying the diverse opportunities through which learners could engage with data practices or skills. Conversely, others might be focused on representation or general inclusion, which shifts attention to ensuring that diverse learners are represented in data science programs, classrooms, or jobs (Li & Koedel, 2017). This framing of equitable outcomes prioritizes expanding access to programs or jobs, but theoretically obscures what experiences might foster or lead to equitable representation.

Our focus on identity development as a learning process fills in the gaps by complementing micro-perspectives from teaching and learning, and having a macro-focus on general access. In the following chapter, we outline a framework for how a learner's emerging occupational identity (Callahan, Ito, Campbell Rea, & Wortman, 2019) as a data scientist could be understood through five components: (1) one's self-positioning, (2) competency beliefs, (3) social capital, (4) structural opportunity, and (5) navigational understanding. Then we shed light on the utility of this identity framework by sharing case vignettes of undergraduate students, who identify with traditionally underrepresented groups and are in a university program that is designed to foster pathways into data science careers. The unique aspect of our case study work is that these students are a self-selected population who are already interested in data science as a field and career pathway. However, within this seemingly ideal student population (who one might expect are of the highest probability to pursue data science), we show how an identity lens helps reveal the complex interplay of social, economic, and political factors that each individual learner grapples with as they both experience the educational program and develop their own evolving decisions about whether they see themselves as data scientists. From these vignettes we illuminate how the *design of educational programs* or *pathway experiences* might attend to these complex factors, complement both instructional and access approaches to equity, and ultimately inform new research hypotheses about how we might promote data science experiences that attend to the social, economic, and political realities of students.

Occupational Identity as a Focal Lens for Data Science Education

A lens of occupational identity as a core construct of interest (Callahan et al., 2019) attunes research directions on some core boundaries. First, we see learners' decisions to pursue and persist – long term – on careers in data science as a focal outcome (versus other valid outcomes such as learning technical data science skills etc.) and this focus often involves different experiences for learners than just those in formal classrooms (Calabrese Barton et al., 2013; Carlone & Johnson, 2007; Polman & Miller, 2010). Second, examining identity necessitates defining what aspects make up one's decision to "be a certain type of person" (Sfard & Prusak, 2005) and how learning environments support or hinder these complex decisions (Ahn, 2019). Identity can be defined in a myriad of ways and the construct is debated and complex. For example, sometimes identity is described as a process of learning to be a part of an established group, such as in a communities of practice framework (Lave & Wenger, 1991). Other times identity is described as the narratives one might tell about oneself (Sfard & Prusack, 2005), a sense of belonging (Jones, Tendhar, & Paretti, 2016) to a group, the ownership a learner might feel (Yip et al., 2014), or perhaps one's imagination of themselves (Holland, Lachicotte, Skinner, & Cain, 2001).

Recognizing the complexity and different layers of identity, we focus our framework on five concepts that are particularly salient for the development of occupational identity, where a learner may develop a deeper sense that they are – for example, a data science person who pursues this identity as a career path. Below, we briefly describe the components that might explain a learner's occupational identity and explain how we conceptualize the components in the context of data science education.

Self-positioning: Imagining my Possible Self in a Field

Learners iteratively evaluate and define possible selves — defined as "cognitive manifestations of goals, aspirations, values, and fears" (Shepard & Marshall, 1999) — in relation to specific situations or goals (Oyserman & Markus, 1990). For example, a learner may position their values, goals, and aspirations to align themselves with possible selves in different careers. This process can motivate them to pursue relevant, career-advancing opportunities (Brown et al., 2020; Markus & Nurins, 1986). In the context of data science education, we might ask learners to reflect on aspects of a future career (e.g., as a data scientist) that are appealing or unappealing to them, and emotions and questions that they have surrounding this career field. This reflection surfaces learners' underlying values around the career and helps them reflect on potential alignment between their goals, aspirations, and values and how they imagine possible selves in the future.

Competency Beliefs: Perceptions of One's Knowledge and Skills

We draw from social cognitive theory (Bandura, 1997; 2001) to capture learners' self-perceptions of their knowledge and skills in the data science field. Competency beliefs can be defined as the extent to which one perceives their ability within a domain to achieve the desired outcomes. One's competency beliefs can influence their behaviors, affects, and motivations, and sustain their commitment to pursuing career-advancing opportunities (Carpi et al., 2017; Eccles & Wigfield, 2002). Treating competency beliefs as an important component of identity development, we might ask learners to identify areas of knowledge and skills that they perceive needing to know, in order to pursue careers as a data scientist, and their perceived competence in these areas. Supporting students' competency beliefs can lead to gains in how students see themselves fitting into science disciplines, promote their confidence, and enhance motivation to learn (Lopatto, 2007; Ryder et al., 1999).

Social Capital: Relationships that One Perceives as Beneficial

Learners pursue goals with the support of social capital — such as a network of families, peers, mentors, and other resources (Bourdieu, 2011). Putnam (2000) describes social capital as bonding and bridging social capital, both of which are necessary to support learners' identity construction. Bonding social capital describes deep, close ties that provide emotional and personal support. They might serve as models that influence how learners self-position themselves and pursue opportunities. Meanwhile, bridging social capital represents less close, but diverse relationships that learners can build on when seeking information and expanded opportunities. Both bonding and bridging social capital can serve as embedded assets, as learners decide on and persist in a career field (Martin et al., 2020). Inviting learners to reflect on (1) their close relationships and (2) mentors in the desired career fields can help them articulate the robust social, emotional, and informational resources that they already possess. It also provides the opportunity to identify gaps in one's social network and broker new relationships, particularly when learners first enter a career field.

Structural Opportunity: Access to Experiences that Learners are Afforded

We further focus on structural opportunities, defined as activities relevant to the pursuit of the desired careers (Azevedo, 2011; Bohnert et al., 2010; McDonnell, 1995). Not all students have access to activities and experiences traditionally valued in data science education, such as coming in with mathematics, statistics, and computer science backgrounds (National Academies of Science, Engineering, and Medicine, 2018). At the same time, students might already engage in data work through informal ways, such as reading, interpreting, and analyzing their own fitness data, or relating data patterns to personal and community experiences (Calabrese Barton et al., 2021; Clegg et al., 2020). It is thus critical to highlight how their everyday experiences and activities are contributing to future careers (Takeuchi et al., 2019). In helping learners to consider structural opportunities that support future careers as data scientists, we ask them to reflect on (1) the opportunities to practice skills or qualities that are connected to the specific career, and (2) everyday experiences that can also be relevant to the career. These questions help us to understand opportunities that learners already have, while also elevating the everyday, informal experiences they already engage in.

Navigational Understanding: Knowledge to Take Advantage of Opportunities and Engage with Organizations or Institutions

Learners also need to navigate multiple layers of requirements in their pathways to a career, from academic (e.g., knowing how to get into college, or completing major-related prerequisites etc.) to career-specific milestones (e.g., obtaining an internship, getting a certain work certification etc.). Learners need to identify the steps they need to take to progress toward a career, reflect on existing experiences, values, and interests, and plan logistically for how they might pursue next steps. One may have a broad concept of their future career goals, but identifying concrete next steps that map onto specific time frames can be challenging for many learners (Brown et al., 2020). Thus, we invite learners to concretely explore (1) the challenges, requirements, and opportunities that exist to get them to their desired careers as data scientists, and (2) the specific steps they are taking within the short and long term. The goal of this exercise is for learners to continually reflect on their navigation and develop a coherent path of action over time, rather than dictating a single correct path.

Observing this Identity Framework in Action

In our work, we are exploring how undergraduate students are developing their occupational identities as they participate in a specialized training and mentorship program in data science. The data science program is housed in a research-intensive, public university that serves a large population of low-income students, first generation college students, and is a designated Hispanic Serving Institution and Asian American, Native American and Pacific Islander Serving Institution. The program is selective and competitive, where each fall, undergraduate students across the whole campus are invited to apply to participate. A cohort of 15 students (maximum) are chosen each year to participate in a year-long fellowship that lasts from January-December of the student's 3rd and 4th year of their undergraduate studies. Thus, students typically start the program at the end of their junior year and complete the fellowship as they are applying to data science graduate programs in the beginning of their senior year of undergraduate study.

The program is designed to support students who voice an interest in pursuing a career in learning analytics or data science and attending graduate school. Priority is given to students who identify with traditionally underrepresented populations, including students from minority racial/ethnic backgrounds (Latine, Black, Indigenous, Pacific Islander), first-generation college students, disabled students, and students from economically disadvantaged backgrounds. During the year-long program, fellows propose and undertake a data science research project under the guidance of a faculty mentor. Fellows are also encouraged to participate in the faculty mentor's research lab or group to gain experience in the cultural and social aspects of working in university research settings. Fellows receive additional training through bi-weekly professional development sessions where they participate in R studio workshops and research seminars, where they learn skills in using R and conducting research. During the summer months, fellows participate in their research labs full time and begin preparing for graduate school applications. The program financially supports fellows through monthly stipends, room and board during the summer session, and by paying for GRE test preparation and graduate school applications.

Of the 14 fellows in the program – who also participate in our ongoing research study – 13 identify as women, and one as a man. In terms of racial/ethnic identity: four fellows identify as Hispanic/Latine, four as Southeast Asian, five as East Asian, and one as White. Six of the fellows shared that they are first-generation college students and one identifies as being disabled. To gain deep understanding of the five components of occupational identity trajectories of our fellows, and how the different components shift and transform over time, we observe the seminars they attend, and schedule a series of interviews with our fellows throughout their fellowship year. During the seminars we produce field notes and analytical memos, paying special attention to the interactions and methods of facilitation at play. At the time of writing this chapter, 18 weeks into the program, we have conducted two interviews with each fellow, one just after fellows have been onboarded and integrated into their research labs (three to four weeks into the program) and the second after fellows have identified a potential individual project to pursue (ten to eleven weeks into the program). Interviews follow a semi-structured interview protocol with items that help us understand each fellow's beliefs and perceptions about their self positioning, competency beliefs, social capital perceptions, structural opportunities, and navigational knowledge. Three more interviews are scheduled for the remainder of the program. Across these interviews, our research aim is to trace the evolution of how the data science fellows put together these different identity facets in their own narratives.

In the following section, we shed light on the affordances of tracing the students' learning through the lens of these identity facets. Specifically, we show how the impactful moments that different fellows remember, perceive, and share with us throughout the program can be observed as a complex intertwining of the five elements of our framework. This understanding then helps us as data science education researchers – and educators or mentors – to reflect on how different aspects of this university program interact with each individual student in variable ways. To illustrate the utility of this analysis, we offer two example vignettes in this chapter. The first focuses on how our fellows are positioning themselves in relation to their imaginations of data science and the second hones in on another set of students' experiences around navigating data science within our current social, political, and economic context.

"It seems more like number crunching" – The Nuances of Imagining One's Self Positioning

One theme that has consistently exhibited itself through the fellows' interviews involves the ways in which their imaginations of themselves as data scientists are intertwined with their competency beliefs, social networks, structural opportunities, navigational awareness, and past histories. Fellows described their perceptions of data science as a field to be emerging and filled with both possibilities and uncertainties. This lack of clarity and stability was related to several fellows describing difficulty in imagining what life would be like as a data science professional.

For example, one fellow Carly (pseudonym) responded in this way when asked about what obstacles she perceives in becoming a data scientist:

> "I don't really know much about data science. I'm not going to lie and how the field actually is, I know it's a bit of research. But to be honest, it seems more like number crunching. And, like examining numbers, so that's what it kind of gets, like, in my definition."

Like Carly, many fellows described their perception of data science, and their role in it, **in terms of technical skills or competencies** they believed to be important (e.g. number crunching). It was helpful for our research team to be able to observe and contextualize these student perceptions – and their focus on technical skills (or competency beliefs) as a primary marker of their imagination of the field – with the outset of the program. Placed within this sociohistorical context, it may make sense that students would focus on competency and skills in the absence of clear understandings of the other aspects that may go into becoming a data science professional.

In the beginning stages of the program, students' prior knowledge and skills were related to their initial self positioning narratives. For example, many of the first generation students in the program voiced initial difficulty in envisioning what life would be like as a data scientist or how data science could help them achieve their goals (which again, is not surprising for students beginning a program). However, our focus on understanding the identity trajectories of the students, enabled us to easily observe how these experiences from first-generation students were in juxtaposition to Gina (pseudonym), who shared that she came to the program already having prior research experience and is the daughter of a professor. When asked about what drew her to a data science program, Gina shared:

> "I'm really excited to actually learn how to use R stuff, because I, like I said before, I've been in a research group for two and a half years, I understand that's something that they really need to use. And I'm excited because I know that'll be good for graduate school prep wise and stuff. I want to get a PhD in education."

Gina came into the program with both **social capital** (daughter of a professor, being comfortable in a research group already) and **structural opportunities** in the past (understanding some tools and skills needed to participate in the research group).

What was illuminating at this stage of the program and of learning about the students, was the **interaction** between the students' self positioning (how they imagined themselves in relation to the field), their initial focus on technical skills as markers of their belonging, and how prior structural opportunities to be comfortable with technical markers (e.g. learning R) already gave some clarity for certain students (Gina) while other students experienced more vagueness and uncertainty. A key point here is that all of the students in this specialized data science program were capable and interested in learning technical skills and tools, but they differed in how their competency beliefs and future imaginations intertwined in these crucial beginning stages of their learning experience.

We also observed how students began with unclear and abstract notions of data science as a field, and along with their anticipation that the field is fast-changing and unpredictable, made it difficult for students to position themselves or know how to see themselves as succeeding in the future. This perception compelled several students to describe **navigational knowledge** and **structural opportunities** as key areas that they were anxious about at the early stages of the program. For example, some students expressed a desire "to see data scientists working in the field and the obstacles that they have while working" and "getting experience doing specific work in the field" in order to gain practical experience for how to manage and navigate their career path.

Still, other learners in the data science cohort described how their different past experiences related to their perceptions of their own positioning in the field. This variation in backgrounds influenced the students' perception of their potential success, interest in the program, and their imagination as a future data scientist. For example, Emily (pseudonym) shared her worries about how her lack of past experiences related to her initial feelings of not belonging to her research group:

> "... since, as I mentioned earlier, it's very new to me. I haven't had the opportunity to take public education classes. So a lot of topics that my lab is working on explaining can be all new to me. So sometimes I feel like I'm left like little behind from others in my lab group, so that sometimes, I kind of feel like maybe kind of don't belong in that group."

Here Emily is describing her perception that her lack of experience in public education puts her at a disadvantage in her research group – which focuses on studying urban students' learning experiences – because she requires more explanation to stay up to speed with her peers (who she perceives as already knowing tacit knowledge she does not have). These initial perceptions, whether accurate or not, contribute to the self positioning narratives that are already forming in the beginning stages of her learning experience.

We offer a final example of how elements of past experiences, competency beliefs, and self positioning come together in entirely different configurations for other learners. For example, Charlotte (pseudonym) was another data science fellow who came into the program with clearly defined research interests and research experience. Her past research experiences led her to have a clear interest in using data science to study issues of K-12 teacher retention, and was also motivated by her past experiences having a "revolving door" of math teachers in her own schooling. While these past experiences inspired Charlotte to have clarity in her purpose for joining the data science program, she also attributed her K-12 experience as giving her a weak background in mathematics (her own competency belief) which she worried might hinder her ability to thrive in the data science field. It was illuminating to observe how the same past experience provided both a positive inspiration for Charlotte's participation in the data science program, and simultaneously afforded a potential anxiety or obstacle in her imagination of herself in the field. Charlotte came into the program with clear research interests (a potential positive driver for her self positioning), but early on in the program experienced difficulties in finding a faculty mentor, research lab, and datasets that exactly matched her well-formed research goals, which led her to worry that she may "run out of steam" and not persist in the program.

"That pressure impacts some of my decisions" – Financial and Social Mobility and their Influence on Identity Stories

A second theme in our initial analysis of fellows at the beginning of this data science program, is that common career decisions about financial and social mobility can be understood in terms of the identity narratives learners are continuously crafting for themselves as they engage in learning experiences. Several data science fellows expressed perceptions of financial investment and opportunities as key factors in their future imaginations of themselves as potential data scientists. Some fellows imagine data science as a possible safeguard for the financial investment of undergraduate education and as a set of skills and competencies that can help them achieve personal goals like supporting their family.

The rising costs of pursuing an undergraduate degree, specifically the cost of tuition and rising housing costs, was a major concern for many fellows. Some fellows shared in interviews that they commute to campus from several hours away, take out large loans to complete their degrees, or work up to 30 hours a week in jobs such as food service, to afford the costs of their undergraduate education. For example, when asked about the challenges he currently faces navigating his academic and professional career path, Ben (pseudonym) shared:

> "Mostly it's been in terms of like, being able to pay for school, it's been really, like rough trying to afford, like having to take out loans and whatnot to be able to afford to, like stay on campus and keep going to classes. Especially since, like, the interest rate has been so high, like the federal loans. It's like eight percent. It's kind of a crazy amount. And I'm like, Oh, my gosh."

The cost of a four year institution was shocking to Ben, who prior to coming to this university was attending community college and living at home.

Some of the designed aspects of the data science program were well aligned to these potential challenges for students. For example, the stipends granted for participation in the fellowship made the program accessible for many students, both attracting them to the program and providing the financial support they needed to take a risk and spend time exploring data science. Financial concerns often provided space for students to ask about navigational processes at the university institution. In the first days of the data science program, confusion about when and how the stipends would be distributed was of immediate concern for several students. At the conclusion of the first seminar and R workshop training, students asked many questions – not about R and learning skills with the statistical package – but about how the stipend may affect their financial aid and housing eligibility. These types of circumstances weigh heavily on how learners **navigate** their career paths and choose which opportunities to pursue in the short term, and influence their long-term imaginations for where they want to go in their future.

Perceptions about social mobility and worth also played major roles in how students learn to navigate systems and position themselves for the future. For example, 12 of the 14 students in the program were pursuing humanities and social science degrees, traditionally seen as less valuable than STEM degrees. Many expressed worry about the value of their education and saw the data science program as a way to give themselves career flexibility, possibly mitigating the income disparity between STEM fields and their interests. Carly, whose desired major was phased out by the university, exemplifies the anxiety felt by learners unsure if their education and training will be worth it, saying:

> “So that brought a lot of fear to me, like, hey, this school is getting rid of it. Are you calling my degree worthless? So I basically like, hey, I need to switch degrees. I don't know which one I just need to switch.”

Fellows in our program often expressed their desire to seek data science related skills to better position themselves for an uncertain future. They recognized data as becoming increasingly important in many fields and the necessity to master data science practices to potentially achieve their goals. Learners in the program also shared how aspects of social capital and self positioning related to their evolving sense of how they might persist in data science; particularly in sharing how they are not only navigating futures for themselves, but for the relationships they care about. Elena (pseudonym), for example, brought up how her first-generation status and the socioeconomic position of her family made her less likely to take big academic or career related risks:

> "Taking care of my parents and like my younger siblings. So that's another pressure that maybe other students don't have. So that's why like, being first-generation, you have that pressure of like your family, and like, I'm the oldest child, so like, you need to succeed. That pressure impacts some of my decisions. And I feel like that could be a challenge in like the future as well. Maybe I'm not as seeking as I would be otherwise."

Elena's first generation status and the responsibilities she has to her family influence what opportunities she seeks and how she navigates her future. Her **social capital** has a direct bearing on the kinds of **structural opportunities** that she sees as safe, valuable, and worth pursuing.

Similarly, Carly's plans for the future are driven by a desire to support her parents and help them get out of debt.

> "So it's, like, stressful to, like, have all these like considerations like, location, income, is this job, you know kind of growing on the market is not growing on the market. And it's just like, those types of like, hesitancies that actually are kind of rooted in like, money, to be honest. That is, like, kind of, like, always gives me hesitancy when I try to like, figure out a career. Right. It's like, do something that has passion, but like, if I have passion, I can't really take care of people that I want to take care of."

Learners today come to data science within our current sociopolitical context, and the pressures from that context influence how they **navigate** their career pathways. The tension Carly experiences between personal interests she brings to the program, her nascent interest in data science, and her desire to provide for her family encapsulates the complex, future imagining done by many learners today.

Discussion and Implications

A central aim of this current chapter is to demonstrate the utility and potential for understanding data science education through a lens of identity. We see an identity lens as a potential way to fill in the gaps, and move away from potentially limiting theoretical framings of data science education and equity. For example, a common approach to designing for equity in a data science education program would provide learners with data and projects relevant to their experience and local community. More engaged in data science work due to the personally relevant presentation of activities, a learner in such a program develops new data science skills and competencies. In some programs, learners may even be contributing to civic or social justice initiatives in their communities as data practitioners (cite). Other programs may be designed to achieve the equitable outcome of increasing diversity and access within the data science workforce (cite). A learner in a program like this may interact with data science professionals and learn about how data science professionals work. The data science professionals and learners may share experiences or aspects of social identity, hopefully helping underrepresented learners envision a space for themselves in the field of data science. We do not present identity development as an alternative or superior equitable outcome, but as a complementary way to conceptualize and assess equitable outcomes in data science education.

The design implications when considering identity development as an equitable outcome needs additional contributions from the field. In our program, situated within a university context with participants who had a preconceived interest in data science or learning analytics, economic conditions were a key concern for learners. We are able to adapt key designs of the program, including professional development, programming, and funding allocation to the evolving occupational identities of our learners. For example, funding was allocated toward GRE test preparation once it became clear that graduate school applications, and the costs associated with them, were hindering participants' navigational understanding of graduate school and how graduate school would fit along their professional trajectory. In K-12 settings, or contexts where learners don't self-select into a data science program, taking an identity lens may elucidate social or pedagogical injustices that inhibit the development of different components of a learners occupational identity.

The difficulty of designing programs for identity development lies in its uniqueness to the person as it is multifaceted and requires attention to various aspects of the learner. The issues of self-positioning and imagining the self in a career choice is not a one-step and done but rather a developmental process that takes time to cultivate. Developing competency beliefs happens over time and necessitates careful reflection on self progress and perceptions of the field. Expanding social capital, especially for marginalized learners, can be a difficult and non-linear process. Developing navigational understanding and pursuing relevant structural opportunities looks different for learners at different age groups. What is clear is that a data science education program that tends to the learner's identity development first must learn how learners position themselves in relation to data science. Providing opportunities for participants to reflect on their past and present experiences and conception of data science is a key step toward understanding data science occupational identity and positions a data science education program to address barriers to realizing an identity as a data scientist.

Scratch Notes:

From our original abstract:

In this chapter, we report on the initial year of a campus-wide, fellowship program to support minoritized undergraduate students learn about the field of educational data science, develop experiences in data science projects, and learn about aspects of the "hidden curriculum" of higher education as they ponder whether they might pursue advanced degrees in the future. The program takes place at a major, state university that is a designated Hispanic-Serving Institution (HSI) and Asian American and Native American Pacific Islander-Serving Institution (AANAPISI), and where nearly 40% of students are Pell Grant recipients (from low-income backgrounds). As PIs of this fellowship program, we are exploring the varied pathways that students -- from diverse backgrounds and majors across campus -- undertake as they decide to enter into the fellowship and experience the program, as well as following them through subsequent years to understand how their evolving perceptions of self and occupational identity evolve after the program. This chapter will analyze in-depth interviews with 15 undergraduate students in the inaugural cohort of the fellowship and take an **identity framework** to shed light on how different students come into a data science experience: with a specific analysis of how their perceptions of past, present, and future selves interact with the designed or intended elements of the fellowship program. In the process, we illuminate how educational program designs can attend to the identity-development pathways of learners, moving beyond only focusing on skills, literacies, or content knowledge aspects of data science.

[Some random thoughts section not sure where this could go, seems like a nice idea to put into implications/conclusion]:

The issue of positioning and imagining the self in a career choice is not a one-step and done but rather a developmental process that takes time to cultivate. The difficulty lies in its uniqueness to the person as it is multifaceted and requires attention to various aspects (relating to identity including emotions). A student shadowing a data scientist for a day may better understand the scope of responsibilities in the role but it does not guarantee that student will immediately see themselves as a data scientist in the near future.

References

Ahn, J. (2019). Drawing inspiration for learning experience design (LX) from diverse perspectives. *The Emerging Learning Design Journal,* 6(1), Article 1. Available at: https://digitalcommons.montclair.edu/eldj/vol6/iss1/1.

Azevedo, F. S. (2011). Lines of practice: A practice-centered theory of interest relationships. Cognition and Instruction, 29(2), 147-184.

Bohnert, A., Fredricks, J., & Randall, E. (2010). Capturing unique dimensions of youth organized activity involvement: Theoretical and methodological considerations. *Review of Educational Research*, *80*(4), 576-610.

Bandura, A. (1997). *Self-efficacy: The exercise of control*. New York: Freeman.

Bandura, A. (2001). Social cognitive theory: An agentic perspective. *Annual Review of Psychology*, *52*(1), 1-26.

Bourdieu, P. (2011). The forms of capital.(1986). *Cultural theory: An anthology*, *1*, 81-93.

Brown et al. (2020). *Striving to Thriving. Occupational identity formation among Black and Hispanic young people and young people from households with lower incomes.* https://www.equitablefutures.org/wp-content/uploads/2020/10/Striving-to-Thriving-Full-Report-October-2020.pdf.

Calabrese Barton, A., Kang, H., Tan, E., O'Neill, T. B., Bautista-Guerra, J., & Brecklin, C. (2013). Crafting a future in science: Tracing middle school girls' identity work over time and space. *American Educational Research Journal*, *50*(1), 37–75.

Calabrese Barton, A., Greenberg, D., Turner, C., Riter, D., Perez, M., Tasker, T., ... & Davis, E. A. (2021). Youth critical data practices in the COVID-19 multipandemic. *Aera Open*, *7*, 23328584211041631.

Callahan, J., Ito, M., Campbell Rea, S., and Wortman, A. (2019). Influences on occupational identity in adolescence: A review of research and programs. Irvine, CA: Connected Learning Alliance.

Carlone, H. B., & Johnson, A. (2007). Understanding the science experiences of successful women of color: Science identity as an analytic lens. *Journal of Research in Science Teaching*, *44*(8), 1187–1218.

Carpi, A., Ronan, D. M., Falconer, H. M., & Lents, N. H. (2017). Cultivating minority scientists: Undergraduate research increases self-efficacy and career ambitions for underrepresented students in STEM. *Journal of Research in Science Teaching*, *54*(2), 169-194.

Clegg, T., Greene, D. M., Beard, N., & Brunson, J. (2020, April). Data everyday: Data literacy practices in a Division I college sports context. In *Proceedings of the 2020 CHI conference on human factors in computing systems* (pp. 1-13).

Clegg, T., Cleveland, K., Weight, E., Greene, D., & Elmqvist, N. (2022). Data everyday as community-driven science: Athletes' critical data literacy practices in collegiate sports contexts. *Journal of Research in Science Teaching*. Retrieved on May 2, 2023 from https://doi.org/10.1002/tea.21842.

Davenport, T. H., & Patil, DJ. (July 15, 2022). Is data scientist still the sexiest job of the 21st century? *Harvard Business Review*. Retrieved on May 1, 2023 from https://hbr.org/2022/07/is-data-scientist-still-the-sexiest-job-of-the-21st-century.

Eccles, J. S., & Wigfield, A. (2002). Motivational beliefs, values, and goals. *Annual review of psychology*, *53*(1), 109-132.

Holland, D., Lachicotte Jr, W. S., Skinner, D., & Cain, C. (2001). *Identity and agency in cultural worlds*. Harvard University Press.

Jones, B. D., Tendhar, C., and Paretti, M. C. (2016). The effects of students' course perceptions on their domain identification, motivational beliefs, and goals." *Journal of Career Development, 43*(5), 383–397.

Kim, J. (2016). Who is teaching data: meeting the demand for data professionals. *Journal of Education for Library and Information Science*, *57*(2), 161-173.

Lave, J., and Wenger, E. (1991). *Situated learning: Legitimate peripheral participation.* Cambridge: Cambridge University Press.

Lang, C., Siemens, G., Wise, A., & Gasevic, D. (Eds.). (2017). *Handbook of learning analytics*. SOLAR, Society for Learning Analytics and Research.

Lee, V. R., Wilkerson, M. H., & Lanouette, K. (2021). A call for a humanistic stance toward K–12 data science education. *Educational Researcher*, *50*(9), 664-672.

Li, D., & Koedel, C. (2017). Representation and salary gaps by race-ethnicity and gender at selective public universities. *Educational Researcher*, *46*(7), 343-354.

Lopatto, D (2007). Undergraduate research experiences support science career decisions and active learning. *CBE-Life Sciences Education*, 6, 297–306.

Markus, H., & Nurins, P. (1986). Possible selves. American Psychologist, 41, 954-969.

Martin, J. P., Stefl, S. K., Cain, L. W., & Pfirman, A. L. (2020). Understanding first-generation undergraduate engineering students' entry and persistence through social capital theory. *International Journal of STEM Education*, *7*(1), 1-22.

McDonnell, L. M. (1995). Opportunity to learn as a research concept and a policy instrument. Educational evaluation and policy analysis, 17(3), 305-322.

National Academies of Sciences, Engineering, and Medicine. (2018). *Data science for undergraduates: Opportunities and options*. National Academies Press.

Oyserman, D., & Markus, H. R. (1990). Possible selves and delinquency. *Journal of personality and social psychology*, *59*(1), 112.

Persaud, A. (2021). Key competencies for big data analytics professions: A multimethod study. *Information Technology & People*, *34*(1), 178-203.

Polman, J. L., & Miller, D. (2010). Changing stories: Trajectories of identification among african american youth in a science outreach apprenticeship. *American Educational Research Journal*, *47*(4), 879–918.

Putnam, R. D. (2000). Bowling alone: America's declining social capital. In *Culture and politics* (pp. 223-234). Palgrave Macmillan, New York.

Ryder, J., Leach, J., & Driver, R. (1999). Undergraduate science students' images of science. *Journal of Research in Science Teaching*, 36(2), 201–219.

Shepard, B., & Marshall, A. (1999). Possible selves mapping: Life-career exploration with young adolescents. *Canadian Journal of Counselling and Psychotherapy*, *33*(1).

Siemens, G., & Baker, R. S. D. (2012). Learning analytics and educational data mining: towards communication and collaboration. In *Proceedings of the 2nd international conference on learning analytics and knowledge* (pp. 252-254).

Takeuchi, L., Vaala, S., & Ahn, J. (2019). Learning across Boundaries: How Parents and Teachers Are Bridging Children's Interests. New York, NW: Joan Ganz Cooney Center at Sesame Workshop.

Yip, J., Clegg, T., Ahn, J., Bonsignore, E., Gubbels, M., Rhodes, E., & Lewittes, B. (2014). The role of identity development within tensions in ownership of science learning. Boulder, CO: International Society of the Learning Sciences.