Characterizing How Engineering Undergraduate Students Define and Develop Data Proficiency

Godwyll Aikins
Department of Mechanical and Civil Engineering
Florida Institute of Technology
Melbourne, Florida
gaikins2015@my.fit.edu

Catherine G. P. Berdanier
Department of Mechanical Engineering
Pennsylvania State University
University Park, Pennsylvania
cgb9@psu.edu

Kim-Doang Nguyen
Department of Mechanical and Civil Engineering
Florida Institute of Technology
Melbourne, Florida
knguyen@fit.edu

Abstract—**This work in progress presents current findings from a funded mixed-methods investigation of the relationship between data proficiency and engineering identity among undergraduate students throughout their curriculum. This study aims to understand ways engineering undergraduate students conceptualize data proficiency and develop data skills over time. Through semi-structured interviews with four undergraduate engineering students from different class levels, we examined their understanding of data proficiency and the importance of data skills in engineering practice. The interviews were guided by the *How People Learn* framework, which provided a lens through which to investigate students' attitudes, beliefs, and experiences related to data and data analysis. The findings suggest that students view data proficiency as an important skill for their future careers but differ in their preferences for learning data skills through assignments, projects, or lectures. This research contributes to the understanding of how engineering students define and develop data proficiency, which can inform the design of effective data skills curricula in engineering education**.

I. Introduction

Data is defined as an object, variable, or piece of information that has the perceived capacity to be collected, stored, and identified [1]. Correspondingly, data skills refer to abilities to interpret data, make connections, and turn data into meanings to solve a problem effectively [2]. Data skills comprise a spectrum of skills including foundational techniques; transformation of data into tangible meanings, insights, or intuitions, and advanced skills such as machine learning and artificial intelligence techniques. To be data proficient, students must contextualize, interpret, and manipulate data to meet stakeholder needs. Data, data skills, and data proficiency are critical because data-enabled technologies are in high demand. Transforming the engineering workplace, technology is instrumental in the ongoing radical shift of jobs enabled by automation [3].

Previous literature demonstrated that the development of engineering identity is critical to the retention and career success of engineering students [4]. An individual's motivation to persist is influenced by expected outcomes, perceived value, and projected costs associated with a given task [5]. Given the increasing attention to data skills in the engineering curriculum, investigating the development of data proficiency connected to engineering identity and motivation is essential to continue to recruit and retain a competitive engineering workforce. The overarching objective of our broader ongoing research initiative is to link data skills with engineering identity.

This work is supported by the National Science Foundation under the EEC Grant #2245022

II. Literature Review

Researchers have sought to establish a connection between specific competencies in the formation of engineering thinking, particularly with respect to problem-solving, systems thinking [6], and design thinking [7]. Of particular relevance to this project, recent research driven by the growing need for engineers capable of programming to meet industry demands has also focused on computational thinking [8], examining how students apply their problem-solving skills in mathematical and algorithmic contexts [9]. Some scholars have suggested that incorporating computational skills earlier and more frequently in engineering curricula enhances the learning experience and leads to better engineering outcomes [10]. Previous studies on computational thinking have demonstrated its positive impact on the development of the engineering identity. Therefore, integrating data skills throughout the curriculum also influences the formation of engineering identity.

Being proficient with data is a crucial aspect of an engineering student's education throughout the engineering pipeline since data is ubiquitous in every facet of the field. In laboratory assignments, for instance, students run experiments; record, organize, and visualize data; perform relevant calculations on the data; extract meaning from the data analysis; and compare results with theories taught in class to establish learned knowledge [11]. In design and project-based assignments, students analyze and visualize data on given tasks and design constraints, dimensions, and materials to create a set of conceptual designs before selecting the best design based on the data available [12]. The activities are essential because practicing engineers follow similar processes when performing their tasks [13].

Therefore, we hypothesize that familiarity, competency, and proficiency with data skills are crucial in the enculturation and formation of engineers.

However, literature has not yet established the links between engineering identity, motivation, and data skills. If the links were made explicit, the research as well as these experiences would impact the formation of engineers and inform practice about the time and methods to introduce data skills in the engineering curriculum. Since the field of data literacy is yet emerging, foundational gaps still need to be solved before the more theoretical links are made. For example, the existing literature has not well demonstrated ways undergraduate engineering students conceptualize data proficiency for themselves or identify gaining the skills. To this end, the research questions that this study answers are as follows:

1) How do engineering undergraduate students conceptualize data proficiency?
2) How do engineering undergraduate students identify opportunities to develop data proficiency in their academic trajectory?

The preliminary results that will be presented in this paper will help to inform how engineering educators can meet their students where they are, identify misconceptions about data and data proficiency, and educate a future data-literate engineering workforce.

III. Theoretical Framework

This study will be primarily guided by Bransford's *How People Learn* (HPL) framework [14], which is an established theoretical framework typically used to identify the most effective ways to develop learning environments from a learner-centric viewpoint. The HPL framework (depicted in Figure 1) advises that learning environments should be designed with a focus on the learner, knowledge, assessment, and community. The three key components that must be present for learning to occur are the content or information being learned; the context in which it is being learned; and the learners, including their prior knowledge, experiences, and motivations.

In this study, we will apply the HPL framework in a slightly different way rather than using it to design or assess specific learning environments and utilize the framework to explore whether and in which manner undergraduate engineering students identify with the HPL elements as they reflect on learning data skills. As data proficiency is often a "hidden competency" that is not taught in a single class, we also examine how students develop data skills over time using the components of Bransford's framework as part of our second research question. The results of this study will then suggest potential pedagogical implications.

Because the broader reach of the project focuses on persistent student motivation and engineering identity, our study is more broadly oriented to modern applications of the expectancy-value theory (in which humans make decisions based on expected success, perceived value, and tangible or intangible costs associated with the decision) [5], and the engineering identity theory in alignment with Godwin [4]. The framing of some of our interview questions align with these theories as well.

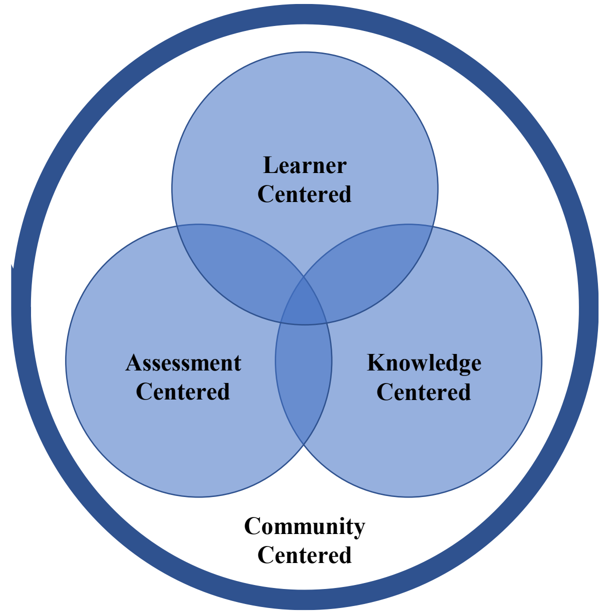

Figure 1. How people learn framework: Creating an effective learning environment. Adapted from [14]

IV. Research Design and Methodology

This work is approved by the Florida Institute of Technology Institutional Review Board as part of a broader project. To address the research questions for this paper, we designed a semi-structured interview protocol informed through the HPL theory. Semi-structured interviews are ideal in their capability of producing rich qualitative data. Qualitative methodologies are commonly employed in discipline-based identity research because interviews allow for easier access to lived experiences and internalized beliefs of identity as they allow participants to describe their experiences and interactions from their own perspectives, which fits well with our investigation into the link between engineering students' identity and data proficiency.

In particular, the interviews were designed to explore the relationship between students' use of data skills and their expectations for success as well as any positive or negative affective memories associated with tasks related to data proficiency. Additionally, attitudes toward the intrinsic value and interest in data skills and their utility value will be examined. For instance, participants will be asked about their affective memories related to learning data skills, how data skills impact their perception of engineering, their willingness to learn advanced data skills, and their motivation to engage data skills in problem-solving. For this work in progress, we report preliminary data from four participants representing the four academic levels of undergraduate engineers. Based on the trial interviews, we refined the interview protocol in the following ways:

- Used broad questions to allow ample detail and a variety of answers without guiding the participant to a particular response.
- Restructured the interview protocol to contain a mix of more- and less-structured questions.
- Made the questions more open-ended and flexible.

- Constructed follow-up questions based on anticipated responses.

The initial interview prompted us to be more flexible in the interview process, which allowed us to explore the respondents' perceptions and follow up on new ideas and themes as they were presented. The revisions are supported by Merriam [15], who suggested that allowing for some unstructured discussion can be valuable, even in semi-structured interviews, as the respondents have the opportunity to raise subtopics that the interviewers may not have considered and to offer fresh insights into the interviewers' topic of interest. The resulting interview protocol from the initial round of interviews is shown in Table 1.

In ongoing work, a screening and interest survey will be distributed to all undergraduate engineering students at the Florida Institute of Technology, a research institution in the southeastern United States. The interest survey will collect information about demographics, including educational background, chosen engineering discipline, familiarity with data skills, and interest in participation. From the interested participants, we will use stratified maximum variation sampling to select a total of n = 40 engineering undergraduate students (n = 10 at each of the first-year, sophomore, junior, and senior levels) while also representing diversity in gender and race of participants. All participants will be provided with an IRB informed consent form. Interviews will be transcribed through a professional transcription service, cleaned for accuracy, and analyzed using directed content analysis methods using a coding framework developed from prior literature. The textual data from the interview transcripts will be analyzed using an abductive approach [16], with the HPL framework as an *a priori* codebook, while allowing the data also to elicit new themes or subthemes to capture the nuances in our data best and comprehend the manner in which the dimensions manifest in the context of data skill development.

V. Preliminary Results

Using the semi-structured interview protocol grounded in the framework established, the work in progress includes a trial round of interviews to refine the interview questions and gain an understanding of the themes associated with the two research. We conducted interviews with four undergraduate engineering students, one from each class (first-year student, sophomore, junior, and senior).

The trial interviews also presented us with some interesting themes. We inquired about the interviewees' definition of a person with data skills or proficiency. Following the question, we asked if they regarded themselves as data proficient. The students defined an individual who is data proficient as someone possessing a specific set of skills. Furthermore, they discussed the various skills, knowledge, abilities, and learning orientations required to achieve data proficiency. During the trial interviews, three main themes emerged. First, upper-class students showed a greater understanding of data proficiency, which was closer to the literature findings compared to first-year students and sophomore students. As a result, the upper-class students viewed all forms of assessment as opportunities to develop their data skills. Last, participants identified that projects are most effective in developing data proficiency.

First Theme: Specificity of Definition The analysis of the interview responses revealed an emergent theme related to the participants' understanding of data proficiency. When asked to define someone who is data proficient, the third-year participant provided a more comprehensive response: "Someone who is data proficient is someone who is able to draw meaningful conclusions from data to understand or solve a problem." In contrast, the lower-class students offered more simplistic responses. Without delving into further details, they described a data-proficient individual as someone who works with the data given to them. Furthermore, when prompted for examples of data skills, the upper-class students demonstrated a higher level of specificity, providing detailed and specific examples of data skills they possess.

Second Theme: Demonstrated Mastery of Data Skills The second theme that emerged from the interviews was closely tied to the upper-class students' deeper understanding of data skills. In contrast, the first-year and sophomore participants primarily associated projects and labs as the primary means of acquiring data skills. However, the upper-class students recognized that all forms of assessments played a role in becoming more data proficient as they acknowledged the importance of experience in skill development.

One senior participant provided an illustrative example: "One of the first things they teach you about school is to look at your answers and make sure they make sense. If you are estimating the force that a human can exert and you come up with 50 kN, something is wrong. It is part of the experience of knowing how much you would expect in terms of magnitude. It is about having seen something similar in the past and knowing what to expect." This insight reflects the senior's understanding that data proficiency encompasses not only technical skills but also the intuition and contextual understanding necessary for meaningful data analysis.

Third Theme: Preference for Authentic Learning The participants consistently expressed a strong preference for projects as the most effective form of assessment for enhancing their data skills. The sentiment was best articulated by one participant who emphasized the role of intuition in the school experience, particularly as an engineering student engaged in practical activities: "I believe that intuition plays a significant role in my school experience, particularly as an engineering student who engages in practical activities. The capstone projects provided me with the opportunity to apply my skills in a tangible, real-world context. However, in most of my other classes, there is not much room for intuition. It is more about following the instructions and applying what we have been taught." Another student highlighted the unique value placed on projects as a form of assessment as they provide a platform for the development and application of data skills in a way that aligns with the practical demands of the engineering field.

VI. Conclusions and Future Work

The preliminary results of this project indicate that engineering undergraduate students view data proficiency as a crucial skill for their future careers and understand its importance in making informed decisions in engineering practice. However, the ways in which students prefer to learn data skills varies.

Through the lens of the HPL framework, this study provides insights into students' attitudes, beliefs, and experiences related to data, which informs the development of effective data science curricula in engineering education. Furthermore, the study highlights the value of incorporating both structured and unstructured components in semi-structured interviews as the approach uncovers new and valuable insights from respondents. The findings from this study contribute to the broader conversation about data proficiency in engineering education. The research offers potential strategies for improving the design and implementation of data science courses and curricula in undergraduate engineering programs.

TABLE 1: INTERVIEW PROTOCOL AND THE CORRESPONDING CONSTRUCTS

Theoretical Constructs	Questions

Learner-Centered: *Pay careful attention to the knowledge, skills, attitudes, and beliefs that learners bring to the educational setting.*	• Tell me about your trajectory to and through engineering.
	  ○ Why did you consider an engineering career?
	  ○ Was anyone instrumental in your decision?
	  ○ What are some of the majors that you considered?
	• What are your future aspirations in engineering?
	• Do you see yourself as someone with high data proficiency?
	• How much do graphs and diagrams help you in understanding a problem or making sense of an answer?
	• What tools have you used in the past to create visualizations?

Knowledge-Centered: *Evaluate specific knowledge that students believe they will acquire and how they improve their data skills.*	• What skills and competencies do you perceive are essential in your professional engineering domain?
	• What comes to mind when you hear the terms *data proficiency* or *data skills*?
	• How would you define someone who has data skills or is data proficient?
	  ○ Give me an example of someone using data skills.
	  ○ How do you overcome not understanding a problem?

Assessment-Centered: *Examine the assessments that students perceive as helpful in enhancing and refining their data skills.*	• Which courses have contributed to your data proficiency?
	• Does completing data skills homework help reinforce the concepts and skills taught in class??
	• Can you describe how to solve an engineering problem once given to you?
	• How have lectures contributed to enhancing your proficiency in handling data?
	• Can you recall a lecture that significantly contributed to your understanding of a topic that was previously unclear to you??
	  ○ What was effective about the methodology the professor used to explain the topic?
	• How do you balance the information presented in a lecture with other resources, such as textbooks or online materials?
	• What types of homework assignments have helped you develop data skills?
	• Have you ever completed a data-intensive project?
	  ○ How did the project compare to homework assignments in terms of helping you learn the material? Did you find the project more engaging or challenging?
	• What are the benefits and drawbacks of learning data skills through homework assignments, projects, or group work?
	  ○ Which method do you think is the most effective for you?

Community-Centered: *Investigate whether students acquire data skills more effectively by learning from or teaching their peers*	• How do your peers and instructors view you as an engineering student?
	• Did you ever work collaboratively on assignments that required you to use data skills? If so, how did the collaboration compare to individual work?
	  ○ What do you think about group projects in general?
	  ○ Do you think group projects alleviate your shortcomings?
	• What roles do you normally take in a group project?

REFERENCES

[1] R. Bhargava, E. Deahl, E. Letouze, A. Noonan, D. Sangokoya, and N. Shoup, *Beyond data literacy: Reinventing community engagement and empowerment in the age of data*. Data-Pop Alliance, 2015.

[2] E. National Academies of Sciences, Medicine *et al.*, *Data science for undergraduates: Opportunities and options*. National Academies Press, 2018.

[3] D. Acemoglu and P. Restrepo, “The race between man and machine: Implications of technology for growth, factor shares, and employment,” *American Economic Review*, vol. 108, no. 6, pp. 1488–1542, June 2018.

[4] A. Godwin, “The development of a measure of engineering identity,” *ASEE Annual Conference & Exposition,* 2016.

[5] J. S. Eccles and A. Wigfield, “Motivational beliefs, values, and goals,” *Annual review of Psychology*, vol. 53, no. 1, pp. 109–132, 2002.

[6] E. Stirgus, M. Nagahi, J. Ma, R. Jaradat, L. Strawderman and D. K. Eakin, “Determinants of Systems Thinking in College Engineering Students: Research Initiation,” 2019 ASEE Annual Conference & Exposition, Tampa, Florida, 2019C. B. Zoltowski, S. M. Eddington, A. O. Brightman, P. M. Buzzanell, and R. Joshi, “Exploring diversity and inclusion in the professional formation of engineers through design sessions,” in *2018 IEEE Frontiers in Education Conference (FIE)*, 2018, pp. 1–5.

[7] H. Y. Durak and M. Saritepeci, “Analysis of the relation between computational thinking skills and various variables with the structural equation model,” *Computers & Education*, vol. 116, pp. 191–202, 2018.

[8] M. A. Hutchison, D. K. Follman, M. Sumpter, and G. M. Bodner, “Factors influencing the self-efficacy beliefs of first-year engineering students,” *Journal of Engineering Education*, vol. 95, no. 1, pp. 39–47, 2006.

[9] A. J. Magana, M. L. Falk, C. Vieira, and M. J. Reese, “A case study of undergraduate engineering students’ computational literacy and self-beliefs about computing in the context of authentic practices,” *Computers in Human Behavior*, vol. 61, pp. 427–442, 2016.

[10] R. Freuler, A. Fentiman, J. Demel, R. J. Gustafson, and J. Merrill, “Developing and implementing hands on laboratory exercises and design projects for first year engineering students,” in *2001 ASEE Annual Conference*, 2001, pp. 6–354.

[11] C. L. Dym, A. M. Agogino, O. Eris, D. D. Frey, and L. J. Leifer, “Engineering design thinking, teaching, and learning,” *Journal of Engineering Education*, vol. 94, no. 1, pp. 103–120, 2005.

[12] D. Jonassen, J. Strobel, and C. B. Lee, “Everyday problem-solving in engineering: Lessons for engineering educators,” *Journal of Engineering Education*, vol. 95, no. 2, pp. 139–151, 2006.

[13] J.D. Bransford, A.L. Brown, R. R. Cocking et al, *How people learn.* Washington, DC: National Academy Press, 2000, vol. 11.

[14] S. B. Merriam and E. J. Tisdell, *Qualitative research: A guide to design and implementation*. John Wiley & Sons, 2015.

[15] I. Tavory and S. Timmermans, *Abductive analysis: Theorizing qualitative research*. University of Chicago Press, 2014.