Capitalizing on a Small Dataset to Refine a Mixed Methods Study

Alison Polasik
Associate Professor of Engineering
Campbell University
Buies Creek, NC, USA
polasik@campbell.edu

***Abstract*— Learning to use computational tools is a critical aspect of many first-year engineering courses. Students in these courses can have varying degrees of prior exposure to computational tools and often exhibit a wide range of intrinsic motivation for learning these skills. Interviews with first-year engineering students were conducted to study the degree to which motivation correlated with intended major within engineering. The pilot round of interviews was conducted in the spring of 2021. This pilot study used the same data set of 8 interviews to progressively refine both the survey and interview collection in the 2021 – 2022 academic year. This paper will review the results of the pilot study and detail the process used to develop and use the finalized protocol and coding schemes for a planned series of 20 interviews in Summer 2022. While this technique developed because of limited data collection during the COVID-19 pandemic, it has been beneficial to the researcher and may provide guidance for future researchers.**

Keywords—Self efficacy, utility value, interviews, computational thinking, first year,

I. Introduction

Engineering students today are “digital natives”, but most engineering educators would agree that learning to use computational tools to model and simulate engineering problems is still cognitively difficult for them. Many engineering students come to their first university courses equipped with a working knowledge of using spreadsheets for data visualization, organization, and calculation – skills that should provide a strong foundation from which to build. Learning to write “scripts” to automate calculations and simulate engineering problems presents a steep hurdle for many students. Faced with this challenge, some students excel and some flounder. While performance in this area is of course a spectrum, this research seeks to study those students for whom difficulty in this critical aspect of engineering skill-building is accompanied by or leads to a decrease in motivational factors such as self-efficacy and utility value.

This works focuses on self-efficacy and utility value related to computational modeling as key aspects of student motivation that are likely to vary between students and may be correlated with students’ intended majors. Self-efficacy is a part of Bandura’s social cognitive theory and is defined as an individual’s judgement of his or her ability to execute a task within a specific domain [1]. In this study, the self-efficacy of engineering students to work with computational tools in general (sometimes also described as “computational thinking” or “programming thinking”) is operationalized by considering their motivation while learning to use MATLAB in their first-year engineering course sequence. Utility value is a sub-construct of expectancy-value theory, which “focuses on the social psychological reasons for people’s choices in achievement settings”. These constructs were chosen as foci of this study because of their established link to learning and performance [2] [3], the availability of validated instruments [4], and observations made in preliminary investigations on the part of the researcher with a population of Materials Science and Engineering students [5].

Several researchers have investigated the epistemology of engineering students specifically related to mathematics and computing, and found differences between computer scientists and engineers, as well as between engineering majors and other students [6] [7] [8]. Thus, it is possible that similar variations could be present between groups of first-year engineering students based on their intended major. Understanding the conditions that lead to lower motivation for learning computational thinking and how to use computational tools is worthwhile because interventions for changing motivation can have more lasting effects on student learning. The research questions for this study are:

1. To what extent do engineering students value computational skills (expectancy value) and how does expectancy value differ across engineering disciplines?
2. To what extent do engineering students think they are able to develop these skills (self-efficacy) and how does self-efficacy differ across engineering disciplines?

This paper outlines one part of a larger mixed methods study that began in Fall of 2020 and will continue until Summer 2023. In 3 subsequent years, first-year engineering students are surveyed (in the fall semester) about their beliefs related to computational methods in engineering. This information is used to recruit and interview a subset of students in the following spring semester. The research design is an explanatory mixed methods study; the hypothesis is that survey data will reveal insight on the research questions, which can then be investigated more deeply with interviews. Preliminary results from the survey have been described elsewhere [9]; these did not provide conclusive answers to the research questions. As such, the role of the interviews in the first year moved to include exploratory elements. Low recruitment response for this this pilot investigation resulted in significant changes to the research design, the method of code development, and ultimately both the survey and interview instruments.

II. Methods

A. The Plan: Data Collection

Participants include students taking an introductory engineering course at two different universities late in the Fall semester. One university (School A) is a large state school with an annual first year engineering enrollment of approximately 1400 students; the second university is a small private university with an annual enrollment of approximately 45 (School B). School A has over a dozen engineering programs (or "majors"), whereas School B has a single engineering major with three concentrations.

At both schools, MATLAB is taught as a key computational tool for engineers in the two-semester sequence for "first year engineering". These first-year students were given a survey that asks students their intended major (which is a key component of the research questions) as well as demographic questions related to confounding factors such as gender, race, ethnicity, first-generation status, and prior experience with computer programming. There are also several questions related to utility value, self-efficacy, self-regulation, and overall opinions related to learning to use MATLAB. Survey administration is timed to occur towards the end of the unit on this topic, after most individual lessons have been completed and as students are beginning to work on the final team project that uses MATLAB. Survey responses were used to recruit a subset of students with whom to conduct interviews. The research was designed as an explanatory mixed-methods project, and it was expected that the interviews would illuminate some of the results from the surveys. Together, these approaches will address the research questions.

A semi-structured interview protocol was developed based on the research questions, initial observations from the survey data, and assumptions about what the finding of the interviews might be. Abbreviated pilot interviews were conducted with second-year engineering students to gauge the clarity of questions and to practice interviewing methods. The interview protocol was also reviewed by the project advisory board.

The general flow of the finalized interview protocol used in Spring 2021 was as follows:

- Context & Background
- Participant Identity
- Participant Experiences with MATLAB
- Participant Beliefs
- Synthesis and Conclusion

Table *1* was used by the interviewer to ensure that each interviewee addressed all key themes.

This material is based upon work funded by the National Science Foundation under Grant No. 2025093 and the PFE: RIEF program.

Table 1: This table was used to guide the interviewer to ensure that all topics were addressed in the interview by checking of each of the 9 boxes during the interview.

Computational Modeling & Engineering	**Computational Modeling & Engineering Subdiscipline**	**Computational Modeling & Self Beliefs**
How will I use computational tools as an engineer?	How will I use computational tools in ________ engineering?	How skilled am I at using computational tools?
How important are computational tools for success in engineering in general?	How important are computational tools for success in __________ engineering?	How important is it that I become skilled at using computational tools?
Specifically with respect to MATLAB	Specifically with respect to MATLAB	Specifically with respect to MATLAB

Interviews were conducted using Zoom from May – June 2021. Interviews were schedule for late in the spring semester of 2021 to allow for initial analysis of the survey data and to allow the students time to reflect on their past course experience. The timing also reflects the difficulties in subject recruitment; though recruitment emails were sent starting in March, many students did not agree to interviews until after the semester had finished. Results were transcribed and deidentified. Pseudonyms for participants were chosen from a list of the most common names worldwide (omitting those that do not use a Latin alphabet). All students in the ultimately small dataset identified as either male or female, and gendered pseudonyms were chosen to reflect interviewee identity.

B. The Plan: Data Analysis

The plan for data analysis was to use an initial skeleton codebook based on the research questions and interview protocol. Open coding [10] would then be used in a subset of pilot interviews to inductively develop detailed codes and a formal codebook. This codebook could then be applied to the entire dataset by the researcher and undergraduate researchers.

C. The Problem

A surprisingly small number of students agreed to be interviewed for this project. The design called for 20 – 25 students combined from both universities; recruitment efforts produced 4 students from each university. Moreover, most interviewees planned to major or specialize in Electrical and Computer Engineering, and interview data from a variety of majors will be required to answer the research questions. Survey responses were similarly lower than expected, and no clear conclusions could be drawn about the relationship between students' intended majors and their motivation with respect to MATLAB from the survey data. Even so, there can be a great deal of information in over eight hours of interviews, and this data was used not only to answer the research questions but also to set the project up for improved data collection methods in the second year.

Three of six steps for qualitative data collection and analysis [11] were iterated through three separate "rounds": engaging with the data, coding extracts, and generating code categories. The first round began with a short a priori set of codes based on the research questions and interview protocol. Round two expanded on the list of codes inductively, and Round 3 refined and tested the resulting codebook and made amendments to the research protocol.

III. Results

The data set of 8 interviews was analyzed over the course of 3 rounds to finely tune both the analysis tools (code book and content analysis design) and the instrument (interview protocol). In each round, all interviews were coded with a slightly different codebook. The overarching goal of this process was to arrive at a detailed codebook that could be used by a team of researchers in subsequent rounds of interviews. Figure *1* gives an overview of the process and shows the refinement of parent codes.

A. Round 1

All interviews were coded using descriptive coding [12] by the primary investigator, and about 30% of the transcribed material was also coded by both a more experienced research mentor and an undergraduate researcher. Some coding was done collaboratively, and all coders were encouraged to make memos in the transcribed text about potential future codes, which were discussed in three separate meetings throughout the process. This process led to the addition of several parent or child codes and identified some potential improvements to the interview protocol. Some key results from this round are:

- All the students operationalized their experiences with using MATLAB in the context of their group project more frequently than they did with the smaller individual assignments. Codes related to group project and their role in it were added.
- The interview protocol did not ask about self-regulation, but several students brought it up. This was added to the code book.
- Interviewees frequently mentioned experiencing enjoyment (or lack thereof) for MATLAB tasks. Questions about this were added to the survey. An open-ended question was also added to the survey: "using three words, describe your experience learning to use MATLAB in (the university's first year engineering course ID)".

B. Round 2

The codebook for round 2 mainly had additions from round 1, and thus contained the largest number of codes. This codebook was applied to 7 of the 8 interviews. At the end of round 2, changes were made to remove or consolidate some codes and detailed descriptions and examples of each code were added to the code book.

- Interviewees did not express a strong relationship between their high school experiences in computer programming (whether curricular or extracurricular) to their engineering identity. Codes related to this potential relationship were replaced with a shorter set of codes to document when students had prior experience (HS or college) rather than the type of experience. The interview protocol was also adjusted to include fewer questions about programming experience before college.
- Memos and conversations between the researchers identified that some nuances related to how the groups dealt with complicated programming tasks were not captured. Two codes were added with the intention that these codes would be used in concert with others. Code *Programming Preference* is used to document a student expressing a preference to do more or less coding (or no preference). Code *Team Dynamics* is used to document different aspects of teams working together and distributing work.

At the conclusion of this round, significant changes were made to the interview protocol. The new protocol focuses much less on prior experience with programming and much more focus is given to the student experiences with the final group project in MATLAB. This protocol is being applied in interviews taking place June – August 2022.

C. Round 3

The finalized code book was developed and applied to the interview data a third and final time in Summer 2022. Some of this coding is occurring in tandem with the second round of interviews, with the intention that this method can inform minor adjustments to clarity or follow-up questions in the interview protocol. Inter-rater reliability analyses are also being conducted using testing tools in Dedoose and will be discussed at the conference.

IV. Discussion and Conclusions

It is worth noting that the most meaningful changes to the research instruments were made *in round 3*, which was not part of the original research design. If the initial data collection in the 2020 – 2021 academic year had gone as planned, much less time would have been spent on perfecting these tools. Knowing that such a small and skewed dataset could not answer the research questions forced the researchers to extract as much meaning and usefulness as possible from what was available, and the research plan for year 2 benefits from this considerably.

This "deep dive" into a data set and the analysis and refinement of the coding method was particularly valuable to the author, who is a novice in engineering education research, and is particularly new to qualitative research. This work is funded under the NSF-RIEF grant, the purpose of which is to "…increase the community of researchers conducting Professional Formation of Engineers (PFE) research." Thus, one major goal of the project is to train the PI – whose education and experience are limited to disciplinary engineering research – to conduct rigorous research in engineering education. Had this project proceeded as designed, the researcher would have spent much time applying and reapplying the research protocol and content analysis methods to many interviews. This could be likened to developing skill at using a microscope to examine a wide variety of specimens. In contrast, the progressive use of refined codebooks can be likened to the researcher learning the details of how the tool itself works and developing expertise in focusing the lenses.

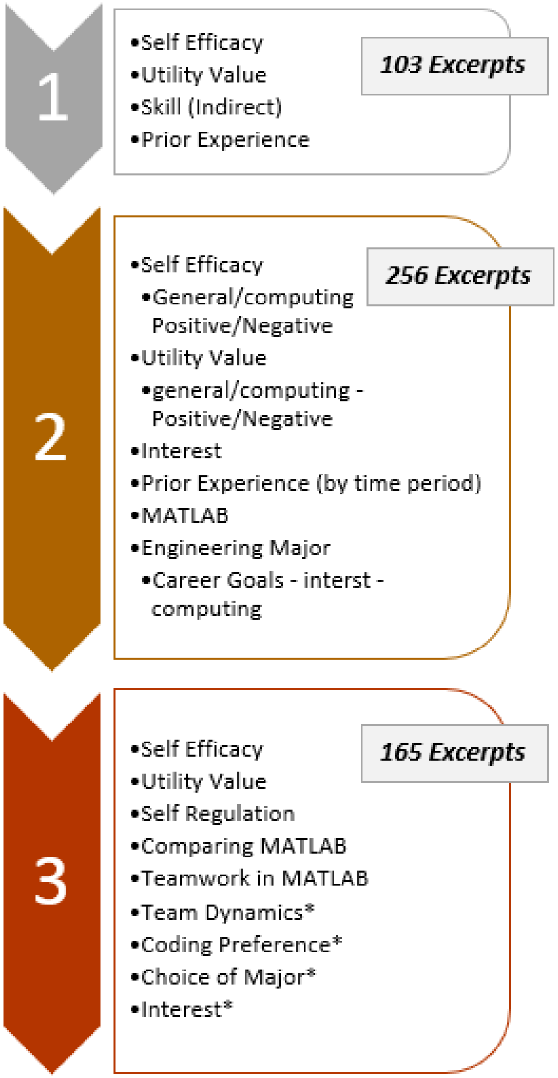

Figure 1: Graphical representation of the 3 coding stages. A new code book is developed at the end of each stage and is applied in the next. Starred codes are intended to be used in several instances as joint codes with others to allow for meaningful analysis of code co-occurrence during the content analysis stage.

As researchers, we know that rigorous qualitative methods can provide insight even with a small number of perspectives. Even so, going through an interview data set three separate times to develop, perfect, and test a finalized codebook before being able to apply it to a full set of data may seem tedious and unnecessary. What began as a response to recruiting challenges has proved immensely valuable to the author, particularly in light of the specific benefits and constraints of the RIEF grant. This method of learning by doing (and doing and doing) made the best use of a foundation in engineering research while learning the research methods of social science. It is hoped that similar techniques could be useful for anyone new to qualitative research in engineering education.

Acknowledgment

The Author thanks Drs. Rachel Kajfez, Alejandra Magana, Julie Martin, and Raeal Moore for their mentorship and guidance.

This material is based upon work supported by the National Science Foundation under Grant No. 2025093 and the PFE: RIEF program. Any opinions, findings, and conclusions or recommendations expressed in this material are those of the author and do not necessarily reflect the views of the National Science Foundation.

References

[1] A. Bandura, "Social foundations of thought and action," Englewood Cliffs, NJ, vol. 1986, (23-28), 1986.

[2] L. Zhang and D. Watkins, "Cognitive development and student approaches to learning: An investigation of Perry's theory with Chinese and U.S. university students," Higher Education, vol. 41, pp. 239-261, 2001. Available: https://www.researchgate.net/publication/227213857_Cognitive_development_and_student_approaches_to_learning_An_investigation_of_Perry%27s_theory_with_Chinese_and_US_university_students. DOI: 10.1023/A:1004151226395.

[3] H. Burn and J. Holloway, "Why should I care? student motivation in an introductory programming course," in 2006 Annual Conference & Exposition, 2006, .

[4] V. Ramalingam and S. Wiedenbeck, "Development and Validation of Scores on a Computer Programming Self-Efficacy Scale and Group Analyses of Novice Programmer Self-Efficacy," Journal of Educational Computing Research, vol. 19, (4), pp. 367-381, 1998. Available: https://journals.sagepub.com/doi/full/10.2190/C670-Y3C8-LTJ1-CT3P. DOI: 10.2190/C670-Y3C8-LTJ1-CT3P.

[5] A. K. Polasik and D. Riegner, "Successes and lessons learned in an undergraduate computational lab sequence for materials science and engineering," in Proceedings of the 2017 ASEE Annual Conference & Exposition, 2017, .

[6] B. A. King and S. Magun-Jackson, "Epistemological beliefs of engineering students," Journal of Technology Studies, vol. 35, (2), pp. 56-64, 2009.

[7] J. Gainsburg, "Engineering students' epistemological views on mathematical methods in engineering," J Eng Educ, vol. 104, (2), pp. 139-166, 2015.

[8] R. M. Marra, B. Palmer and T. A. Litzinger, "The effects of a first-year engineering design course on student intellectual development as measured by the Perry scheme," J Eng Educ, vol. 89, (1), pp. 39-45, 2000.

[9] A. Polasik, A. Suggs and R. Kajfez, "Work in progress: A study of variations in motivation and efficacy for computational modeling in first-year engineering students," in 2021 IEEE Frontiers in Education Conference (FIE), 2021, .

[10] M. Q. Patton, Qualitative Research & Evaluation Methods. 2002.

[11] K. L. Peel, "A beginner's guide to applied educational research using thematic analysis," Practical Assessment, Research, and Evaluation, vol. 25, (1), pp. 2, 2020.

[12] J. Saldaña, Fundamentals of Qualitative Research. 2011.