ASEE 2022 ANNUAL CONFERENCE
Excellence Through Diversity
MINNEAPOLIS, MINNESOTA, JUNE 26TH-29TH, 2022 ASEE
Paper ID #38260

Changing the Paradigm: Developing a Framework for Secondary Analysis of EER Qualitative Datasets

Holly M Matusovich (Associate Professor)

Dr. Holly Matusovich is the Associate Dean for Graduate and Professional Studies in the College of Engineering at Virginia Tech and a Professor in the Department of Engineering Education where she has also served in key leadership positions. Dr. Matusovich is recognized for her research and leadership related to graduate student mentoring and faculty development. She won the Hokie Supervisor Spotlight Award in 2014, received the College of Engineering Graduate Student Mentor Award in 2018, and was inducted into the Virginia Tech Academy of Faculty Leadership in 2020. Dr. Matusovich has been a PI/Co-PI on 19 funded research projects including the NSF CAREER Award, with her share of funding being nearly $3 million. She has co-authored 2 book chapters, 34 journal publications, and more than 80 conference papers. She is recognized for her research and teaching, including Dean's Awards for Outstanding New Faculty, Outstanding Teacher Award, and a Faculty Fellow. Dr. Matusovich has served the Educational Research and Methods (ERM) division of ASEE in many capacities over the past 10+ years including serving as Chair from 2017-2019. Dr. Matusovich is currently the Editor-in-Chief of the journal, Advances in Engineering Education and she serves on the ASEE committee for Scholarly Publications.

Jennifer M Case (Chair, Engineering Education)

Jennifer Case is Professor and Head of the Department of Engineering Education at Virginia Tech in the USA. Prior to her appointment in this post she was a Professor in the Department of Chemical Engineering at the University of Cape Town, where she retains an honorary appointment. She completed postgraduate studies in the UK, Australia and South Africa. With more than two decades of undergraduate teaching and curriculum reform work, she is a well-regarded researcher in engineering education and higher education. Her work especially on the student experience of learning as well as on topics around teaching and curriculum, has been widely published. She was a founding member of the Centre for Engineering Education (CREE) and served twice as its Director, as well as being the founding president of the South African Society for Engineering Education (SASEE). She is a joint editor-in-chief for the international journal Higher Education.

Marie C. Paretti (Professor)

Marie C. Paretti is a Professor of Engineering Education at Virginia Tech, where she is Associate Director of the Virginia Tech Center for Coastal Studies and Education Director of the interdisciplinary Disaster Resilience and Risk Management graduate program. She received a B.S. in chemical engineering and an M.A. in English from Virginia Tech, and a Ph.D. in English from the University of Wisconsin-Madison. Her research focuses on communication and collaboration, design education, and identity (including race, gender, class, and other demographic identities) in engineering. She was awarded a CAREER grant from the National Science Foundation to study expert teaching in capstone design courses, and she is PI or co-PI on numerous NSF grants exploring communication, teamwork, design, identity, and inclusion in engineering. Drawing on theories of situated learning and identity development, her research explores examines the ways in which engineering education supports students' professional development in a range of contexts across multiple dimensions of identity.

Joachim Walther (Professor)

Changing the Paradigm: Developing a Framework for Secondary Analysis of EER Qualitative Datasets

Abstract

This paper reports on a project funded through the Engineering Education and Centers (EEC) Division of the National Science Foundation. Since 2010, EEC has funded more than 500 proposals totaling over $150 million through engineering education research (EER) programs such as Research in Engineering Education (REE) and Research in the Formation of Engineers (RFE), to enhance understanding and improve practice. The resulting archive of robust qualitative and quantitative data represents a vast untapped potential to exponentially increase the impact of EEC funding and transform engineering education. But tapping this potential has thus far been an intractable problem, despite ongoing calls for data sharing by public funders of research. Changing the paradigm of single-use data collection requires actionable, proven practices for effective, ethical data sharing, coupled with sufficient incentives to both share and use existing data. To that end, this project draws together a team of experts to overcome substantial obstacles in qualitative data sharing by building a framework to guide secondary analysis in engineering education research (EER), and to test this framework using pioneering data sets. Herein, we report on accomplishments within the first year of the project during which time we gathered a group of 13 expert qualitative researchers to engage in the first of a series of working meetings intended to meet our project goals. We came into this first workshop with a potentially limiting definition of secondary data analysis and the idea that people would want to share existing datasets if we could find ways around anticipated hurdles. However, the workshop yielded a broader definition of secondary data analysis and revealed a stronger interest in creating new datasets designed for sharing rather than sharing existing datasets. Thus, we have reconceived our second phase as one that is a cohesive effort based on an inclusive "open cohort model" to pilot projects related to secondary data analysis.

Introduction

Since 2010, EEC has funded more than 500 proposals totaling over $150 million through engineering education research (EER) programs such as Research in Engineering Education (REE) and Research in the Formation of Engineers (RFE), to enhance understanding and improve practice. The resulting archive of robust qualitative and quantitative data represents a vast untapped potential to exponentially increase the impact of EEC funding and transform engineering education. But tapping this potential has thus far been an intractable problem, despite ongoing calls for data sharing by public funders of research. Even after work done by Johri, Madhavan and colleagues in 2016 to identify barriers to and strategies for data reuse in engineering education [1], less than a dozen studies funded by EEC have involved data reuse, almost always with quantitative data. The recent meta-synthesis across higher education by Perrier et al. demonstrates the breadth of the problem, highlighting ubiquitous concerns about "data integrity, responsible conduct of research, feasibility of sharing data, and value of sharing data" [2]. These concerns are compounded by **disciplinary and publication practices** that value original data over integrative efforts based on secondary analysis. Additionally, **institutional reward structures** are based on accounting of individual accomplishments and thus discourage more integrated collaboration implied by broad based data sharing. Finally, **funding priorities** stress the novelty of all aspects of proposed work and thus imply a bias for new data generation.

Changing the paradigm of single-use data collection requires actionable, proven practices for effective, ethical data sharing, coupled with sufficient incentives to both share and use existing data. To that end, this project draws together a team of experts to overcome substantial obstacles in qualitative data sharing by building a framework to guide secondary analysis in engineering education research (EER), and to test this framework using pioneering data sets. By bringing together established and emerging scholars over a two-year period, we seek to deliver a tested framework that outlines effective methodological practices for:

- Sharing data both informally and formally
- Putting datasets in the public domain
- Creating combined datasets
- Performing secondary analyses of both qualitative and quantitative data
- Publishing and disseminating these analyses
- Securing funding to support this work
- Valuing and validating this work within the field

Herein, we report on accomplishments within the first year of the project during which time we gathered a group of 13 expert qualitative researchers to engage in the first of a series of working meetings intended to meet our project goals.

Purpose of the Workshop

Through this first workshop, we aimed to scope out the existing knowledge, as well as identify the key challenges that need to be overcome in moving towards sharing datasets for secondary analysis. At the workshop, participants engaged in conversations driven by experiences to date with secondary analysis (successes and failures) as well as identifying general challenges in gathering and curating qualitative datasets.

Workshop Participants

We invited participants based on criteria that they had published research using a qualitative data set and that the data was generated through a study that was NSF-funded. Additionally, we sought diversity in participants, areas of expertise, types of data, and demographics represented in the dataset. To develop a list of potential participants, a graduate research assistant compiled a list of recent publications from engineering education journals including information such as the title of the work, authors, methodology, data source, number of participants, demographics of the participants, focus of the article, and a determination if the data was already a secondary analysis or not. This list was further refined to identify studies that were funded by NSF and for those studies that were, gathering information on the division of NSF that funded the study, the grant title, grant dates, and the project PI. The project team then reviewed the list of potential participants to identify a sub-group to be invited based on the previously described parameters for diversity.

In terms of demographics, the 13 participants who attended the workshop represented educational institutions of different size and focus (e.g., large and small, MSI and PWI), different educational backgrounds (e.g., engineering, education, engineering education, science, etc.), different faculty ranks (assistant, associate, and full professors, Department Chair, etc.), and different geographic regions of the United States. Participants had also focused on different populations of students and used a variety of qualitative research methods in their studies.

Organization of the Workshop

The workshop was scheduled as 1.5 days of working time. The original agenda is included as Appendix A. Participants arrived the day before and had the option to join a group dinner.

Day 1

The workshop opened with informal socializing and meeting time followed by the formal meeting opening with introductions of participants and the project team and setting expectations for the workshop. The first session sought to gather information about participants' prior experience with secondary data analysis by asking them to work in groups to answer the following three questions:

- Have you conducted data analysis on a secondary qualitative dataset? Yes/No
- Have you previously shared one of your datasets for secondary data analysis? Yes/No
- Would you be open to sharing one or more of your datasets for secondary data analysis? Yes/ No/ Maybe/ I don't know

We then introduced the following two discussion questions:

- From your perspective, what are the most promising opportunities associated with secondary data analysis?
- What are your concerns about secondary data analysis? -

We captured notes on large and small post-it sheets. Understanding participant experiences gave us a common, collective launching point for the day.

Following this discussion time, we asked participants to briefly introduce their datasets on post-it notes that we hung around the room to create a gallery. Participants were encouraged to think of this activity as a short version of a methods section of a conference or journal paper that includes the critical details associated with the mechanics of the data set. Over a coffee break, participants had a chance to circulate the room and explore the different data sets.

Dr. Matt Ohland and Dr. Robin Adams presented (separately) on their experiences with large data sets. These large datasets include the Multiple-Institution Database for Investigating Engineering Longitudinal Development (MIDFIELD) and the Design Thinking Research Symposium. MIDFIELD contains "student record data for all undergraduate, degree-seeking students at partner institutions" https://midfield.online/. DTRS is an intentionally designed data sharing project that was designed to promote transformation of design education [3]. Each talk was again followed by reflection in groups identifying elements that resonated.

Following a lunch break, participants engaged in a second deeper round of introducing their data sets. In this phase, we wanted them to think about the personal elements of and their attachments to the data. The instructions were as follows: "Create a 1st person narrative in which your data set introduces itself to the group. An overview of "who" the data set is could also include what its "character traits" are and how the data set feels about itself. You could also consider traits or achievements the data set is proud, happy, or excited about as well as some of the aspects that make it feel shy, embarrassed, or anxious." Authors then briefly shared with the groups and the room at large.

As a group, we then paused to take stock of our thinking and the discussion points that had been raised so far. This was followed by sharing of existing resources related to secondary data analysis.

We followed this "taking stock" exercise with one final group activity for the day. We had intended to have the groups talk about purposes for doing secondary analysis. However, emergent from the prior discussion was the need to focus on several challenges participants saw with secondary data analysis:

1. Issues of training newer researchers (also involving training into your dataset)
2. How to build collaborations/collaborative communities needed to share data
3. Integrating datasets in a community of researchers who have been researching similar topics (not just a one-way exchange)

Participants chose the topic that most interested them. Following group discussion, we had report outs to the larger group.

The final activity of the day was to revise the plan for the second day considering conversations and outcomes from the first day.

Day 2

Day 2 started with Dr. Lisa Benson responding to what she heard on Day 1 and thinking about it from an editor's perspective and offering insights on publishing using secondary data analysis.

Drawing on the session from Day 1 that a "rapid prototyping session" might be helpful, participants had working time in random pairs to create a process for:

- sharing a dataset that was not designed to be shared (choose one of the existing datasets), or
- designing studies with data intending to be shared

The intended outcomes of the session were to develop process principles for sharing data relative to the specific purpose considered and ways to build community around this idea.

Following this session, we had one final session brainstorming next steps for the project and what the project team needed to do/provide before the second workshop in the series.

Outcomes of the Workshop

We came into the workshop with a potentially limiting definition of secondary data analysis and the idea that people would want to share existing datasets if we could find ways around anticipated hurdles. However, the workshop yielded a broader definition of secondary data analysis and revealed a stronger interest in creating new datasets designed for sharing rather than sharing existing datasets. At the start of the workshop, our loose definition of secondary data analysis was "the analysis of a single dataset that has been previously collected and analyzed by a different researcher/ set of researchers." We saw advantages to secondary data analysis such as resources savings such as time and money associated with data analysis and a way to be better stewards of the data by analyzing it more completely. However, during the workshop we found that participants were less interested in secondary analysis of single datasets and more interested in combining datasets on related topics or collecting new data with the intention of sharing it

broadly. The idea of data as a product in and of itself versus the analysis outcomes as the archived product emerged as particularly salient. Importantly, it was not as much a concern over sharing an existing set as it was a lack of desire to revisit the data set given the various limitations each dataset encompassed given that they were not intentionally designed for sharing.

As suggested in the workshop description, as a team we re-organized the workshop on the fly to address the emergent ideas. We are thankful that we took this approach as it led to a different set of outcomes than we originally intended. While our original concerns around the need for careful thinking about the process and culture associated with data collection and analysis (i.e., IRB approval, concerns over who would get access to the data and what would they do with it, etc.), we also found a strong desire to continue engaging in conversations on secondary analysis. Participants were particularly interested in thinking about how to expand the conversation and invite more people into the discussion and the practice; participants were mindful that this workshop group needed to be intentional about not becoming a gatekeeping group for who has access to or gets invited to engage in secondary analysis.

Next Steps

Initially, our team conceived Year 2 of the project as a way to pair new researchers with existing data sets to test approaches to secondary analysis. In response to the first workshop, we have reconceived our second phase as one that is a cohesive effort based on an inclusive "open cohort model" to pilot projects related to secondary data analysis. We are planning to support three types of projects related to secondary data analysis:

- Planning projects which have expected outcomes that include scoping of a secondary data analysis project and articulating lessons learned with a view to developing best practices.
- Planning + data analysis projects with expected outcomes including scoping of secondary data analysis project, engaging in preliminary data analysis, and draft conference publication or NSF proposal.
- Synthesis projects. These projects may have varied outcomes, but potential ideas include planning/conducting ASEE workshop, writing an editorial, proposing/curating a special issue for a journal, or looking across other funded projects to synthesize outcomes.

This spring, we will convene participants from the original workshop and people they identify as potential contributors to review outcomes from the first workshop and collaboratively develop a coherent set of pilot projects to conduct over the summer. We will use the outcomes from these projects as a foundation to reach out to the broader engineering education community about the affordances of secondary data analysis and best practices for conducting this type of work.

Acknowledgement:

This material is based upon work supported by the National Science Foundation under Grant EEC-8060753.

References

[1] A. Johri, S. Yang, M. Vorvoreanu, and K. Madhavan, "Perceptions and Practices of Data Sharing in Engineering Education," *Advances in Engineering Education,* vol. 5, no. 2, p. n2, 2016.

[2] L. Perrier, E. Blondal, and H. MacDonald, "The views, perspectives, and experiences of academic researchers with data sharing and reuse: A meta-synthesis," *PloS one*, vol. 15, no. 2, p. e0229182, 2020.
[3] R.S. Adams and J. Siddiqui, “Purdue DTRS – Design Review Conversations Database.” XRoads Technical Report, TR-01-13. Purdue University, W. Lafayette, IN, 2014.

Appendix A: Original Workshop Agenda

Workshop 1
4 - 5 October 2021

Participants arrive on Sunday 3 October. There will be the option to join for dinner at the hotel restaurant at 18h00.

Day 1 Monday 4 October 2021

8h30 Coffee and chatting
8h45 Welcome and setting the scene
9h00 Meeting the participants and sharing workshop expectations and initial thoughts and experiences in relation to secondary analysis
9h30 Introducing your dataset
9h45 Coffee break
10h00 Presentation by Dr. Matt Ohland, Purdue University
Talking about the MIDFIELD project - what is working well, what is challenging
10h30 Small group reflections on Matt's talk
11h00 Presentation via zoom from Dr. Robin Adams, Purdue University
Talking about the DTRS project which was designed with data sharing in mind.
11h30 Small group reflections on Robin's talk
12h00 Lunch (box lunch pre-ordered through hopeh5@vt.edu)
13h00 Tell the story of your dataset
14h00 Some resources in relation to secondary analysis and an introduction to our google space
15h00 Purposes for doing secondary analysis - group deliberations
16h00 Planning for Day 2
16h30 End of Day 1

18h00 Option to meet workshop participants for dinner (venue TBC)

Day 2 Tuesday 5 October 2021

8h00 Coffee and chatting
8h30 Presentation by Dr. Lisa Benson, Clemson University
Reflections on the deliberations in Day 1. Insights from her editing experience on publishing from secondary analysis
9h00 Identifying processes needed for secondary analysis - in groups
10h00 Plenary feedback and discussion
10h30 Tea/coffee break
10h45 Plans for next steps in the project and next workshop
11h30 Evaluation
11h45 Confirming processes for reimbursement
12h00 End of Day 2
12h00 Lunch and departures for the airport