

Making the Multiple Institution Database for Investigating Engineering Longitudinal Development (MIDFIELD) More Accessible to Researchers

A Special Session

Matthew W. Ohland, Russell A. Long
Engineering Education
Purdue University
West Lafayette, Indiana, USA
ohland@purdue.edu, ralong@purdue.edu

Richard A. Layton
Mechanical Engineering
Rose-Hulman Institute of Technology
Terre Haute, Indiana, USA
layton@rose-hulman.edu

Susan M. Lord
Electrical Engineering
University of San Diego
San Diego, California, USA
slord@sandiego.edu

Marisa K. Orr
Mechanical Engineering
Louisiana Tech University
San Diego, California, USA
marisao@latech.edu

Catherine E. Brawner
Research Triangle Educational Consultants
Raleigh, North Carolina, USA
Cathy@ResearchTriangleEducationalConsultants.com

Abstract— *The Multiple Institution Database for Investigating Engineering Longitudinal Development (MIDFIELD) is expanding to include 113 institutions and is being redesigned and archived to be more accessible to researchers. This special session will describe how researchers can better use or gain access to MIDFIELD. At the conclusion of the session participants should be able to: describe MIDFIELD including common data elements, discuss how new variables can be derived from MIDFIELD, understand what is necessary to access the data on the Interuniversity Consortium for Political and Social Research, and define quantitative and qualitative data types and structures and outline research questions and methods of personal interest to them.*

Keywords—MIDFIELD, database, student, engineering

I. INTRODUCTION

The Multiple-Institution Database for Investigating Engineering Longitudinal Development (MIDFIELD) provides longitudinal data for 1,023,237 undergraduates at 11 large public research universities since fall 1987 [1]. Of those, 210,725 students ever declared engineering as a major. MIDFIELD includes demographic, enrollment, course performance, and graduation data. MIDFIELD institutions include 7 of the 50 largest U.S. engineering programs in terms of engineering bachelor's degrees awarded, resulting in a population that includes 10% of all undergraduate engineering graduates of U.S. engineering programs. MIDFIELD includes 22% female engineering students, which aligns with national averages of 20% to 25% percent from 1999 to 2013 [2]. African-American students are significantly overrepresented

in the MIDFIELD dataset—partner schools graduate 15% of all US African-American engineering B.S. degree recipients each year, because the MIDFIELD participants include six of the top twenty producers of African-American engineering graduates, including two Historically Black Colleges and Universities (HBCUs). The graduation percentage of Hispanics is not representative of other U.S. programs. Three percent of MIDFIELD engineering bachelor's degrees are awarded to Hispanics while 9% of engineering bachelor's degrees in the U.S. are awarded to Hispanics. Hispanic students are particularly concentrated at two institutions in the database that account for 65 percent of the Hispanic students in MIDFIELD. All other groups are representative of a national sample. Even with these limitations, MIDFIELD is large enough to support analyses by race/ethnicity and gender. In this way, we avoid the limitations of synthetic cohorts or cross-sectional analyses.

MIDFIELD is growing and has been funded by the National Science Foundation to increase the number of partner institutions to 113. Students in the expanded MIDFIELD will comprise over half of the undergraduate engineering degrees awarded at U. S. institutions in any given year during the past 25 years. The expanded MIDFIELD will contain unit record data for almost 10 million individual students. The expanded MIDFIELD will also contain minority serving institutions, and institutions from a broad range of research classifications.

II. ENGAGING THE COMMUNITY DIRECTLY THROUGH WORKSHOPS

Well-designed, interactive workshops are an established technique for disseminating and promoting effective practices. In engineering education, for example, the 22-year-old National Effective Teaching Institute (NETI) has introduced nearly 1200 participants from 236 schools to the principles and practices of active and cooperative learning [3]. Three of the PIs have been invited as Fellows to facilitate that workshop, two of whom are now Co-Directors of the NETI. In data visualization, workshops like those by Jean-luc Doumont, Stephen Few, and Edward Tufte have disseminated and promoted critical thinking about quantitative data and effective communication of the stories data tell [4-6]. Members of our research team have regularly developed and delivered interactive workshops on teaching and student learning, student teaming, and engineering education research for nearly 20 years.

A. The Intended Audiences for this Session

The intended audiences for this session are:

- Researchers who are interested in having access to MIDFIELD in their own research.
- Researchers who are interested in figuring out what advantages there would be for their institution to join the MIDFIELD partnership

Why these two audiences would be interested in the session:

- MIDFIELD is a unique research resource. A plan is in place to share that resource with the larger research community.
- This session represents an opportunity for the MIDFIELD team and interested researchers to co-design a process to make the database as accessible as possible to accelerate its use by the wider community.

B. Rationale for Presenting the Session

Reducing data to “dashboards” for easy comprehension and quick decisions by educational administrators and policy makers results in data that are averaged, aggregated, and simplified, but aggregate summaries are incapable of describing the complexity of pathways comprising the undergraduate experience. The complexity of the experience is reflected in the complexity of the data. The goal of analysis is to discover the meaning in the complexity and communicate it effectively—often requiring effort in data management and expertise in data visualization that goes beyond dashboards and pie charts. Once the quantitative “what” of a story emerges, the qualitative “why” prompts further research. Our goal in this session is to develop and support a research community that expands and critiques our methods and findings, building an evidentiary basis for promoting effective policies in support of undergraduate learning in every discipline.

C. Learning Objectives

At the conclusion of the session participants should be able to: describe MIDFIELD including common data elements, discuss how new variables can be derived from MIDFIELD, understand what is necessary to access the data on the Interuniversity Consortium for Political and Social Research, and define quantitative and qualitative data types and structures and outline research questions and methods of personal interest to them.

D. Interaction During the Session

The session format will include brief talks and demonstrations by the facilitators interleaved with participant interactions including think-pair-share, small-group discussion, and then working in groups arranged by research objective. We will solicit participants’ research interests, connect their interests to methods of analysis, and build a common vocabulary for exploring student unit-record data.

III. DESCRIPTION OF THE SESSION

Our session designs are informed by years of experience in designing and implementing learning experiences for students in classes and labs and colleagues in workshops. We plan activities that engage participants, provide training in relevant methods, and challenge them to change their practice. We design the activities to meet learning objectives and enable every participant to leave the session with a specific and personal plan of action for implementing aspects of the special session relevant to their goals.

A. What is MIDFIELD? (15 minutes with handout plus 5 minutes for questions)

Describe the MIDFIELD database including listing common elements of student unit-record data

B. How do derived variables help in research? (5 minutes with handout support plus 5 minutes for questions / discussion)

Discuss how new variables can be derived from the MIDFIELD data set and used in research. Seek input on what derived variables would be helpful.

C. How will researchers have access to the MIDFIELD data? (10 minutes including questions)

A plan has been developed to archive the data with the Interuniversity Consortium for Political and Social Research, which specializes in handling and sharing large datasets. Two restricted-use data dissemination agreements will be described.

D. What resources would facilitate the use of MIDFIELD?
(40 minutes including breakout, questions, and report out)

As part of the larger project, the MIDFIELD team will host a longer institute and provide sample data files and sample programs to conduct certain common analyses. Participants will consider what research questions interest them and be regrouped according to those interests to identify resources that would facilitate those various research uses.

E. Wrapping up. (10 minutes)

In wrap-up, the facilitators will gather feedback from participants in preparation for the design of research support materials

REFERENCES

- [1] Ohland, M.W., & Long, R.A. (2016, in press). The Multiple-Institution Database for Investigating Engineering Longitudinal Development: an experiential case study of data sharing and reuse. *Advances in Engineering Education*.
- [2] B. L. Yoder, *Engineering by the numbers*, American Society Engineering Education. Washington, DC, USA, 2013. Retrieved from http://www.asee.org/papers-and-publications/publications/14_11-47.pdf
- [3] ASEE, National Effective Teaching Institute (NETI): American Society for Engineering Education, <https://www.asee.org/conferences-and-events/conferences/neti>.
- [4] Workshops by Stephen Few, <http://www.perceptualedge.com/workshops.php>.
- [5] Workshops by Jean-luc Doumont, <http://www.principiae.be/X0200.php>.
- [6] Workshops by Edward Tufte, <http://www.edwardtufte.com/tufte/courses>.