Expanding Access to MIDFIELD: Strategies for Sharing Data Infrastructure for Research

Susan M. Lord
Integrated Engineering
University of San Diego
San Diego, CA USA
slord@sandigo.edu

Marisa K. Orr
Engineering and Science Education
Mechanical Engineering
Clemson University
Clemson, SC USA
marisak@clemson.edu

M.W. Ohland, R. A. Long, H. Ebrahiminejad, and H. Al Yagoub
School of Engineering Education
Purdue University
West Lafayette, IN USA
ohland@purdue.edu, ralong@purdue.edu,
hebrahim@purdue.edu, halyagou@purdue.edu

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

Abstract—**This work-in-progress presents an overview of current research with the Multiple-Institution Database for Investigating Longitudinal Development (MIDFIELD) and strategies for inviting and supporting other researchers to use this database in their own research. MIDFIELD is a resource for the study of students that includes longitudinal, de-identified, whole population data for multiple institutions. This enables researchers to examine student characteristics (such as race/ethnicity, sex, socioeconomic indicators) and curricular pathways (including coursework) by institution and over time. Because the dataset contains records of all students matriculating over a period of time, researchers can study students across all disciplines, not just engineering.**

The MIDFIELD team aims to educate the broader research community, expand the network of researchers capable of conducting this research, and share innovative research methods in addition to the actual data. We have offered workshops at several conferences as well as the MIDFIELD Institute. The inaugural MIDFIELD Institute brought together researchers from across the USA for two days in 2019. In an evaluation at the end of the Institute, participants were pleased with the Institute and rated it highly. There was high agreement about the value of the content and delivery. Most participants reported that the instituted lived up to their expectations and was a good way to learn about how to use MIDFIELD data. Areas of improvement suggested by participants included making it longer and more challenging. Several of the MIDFIELD Institute attendees are presenting works-in-progress together at FIE 2020 to provide attendees with a range of examples of research that is possible with MIDFIELD.

Keywords—retention, pathways, faculty development

I. INTRODUCTION

This work-in-progress presents an overview of current research with the Multiple-Institution Database for Investigating Longitudinal Development (MIDFIELD) and strategies that have been initiated for inviting and supporting other researchers to use this database in their own research. MIDFIELD is a resource for the study of students that includes longitudinal, whole population data for multiple institutions. This enables researchers to examine student characteristics (race/ethnicity, sex, prior achievement, socioeconomic indicators) and curricular pathways (including coursework) by institution and over time. Because the dataset contains students records of all students matriculating over a period of time, researchers can study students across all disciplines, not just engineering. More details about the dataset, including data security and confidentiality available in [1]. A comparison of MIDFIELD to a national dataset in the USA is available in [2].

II. IMPACT OF MIDFIELD RESEARCH

Research with high impact in the engineering education community has been conducted with MIDFIELD. For example, MIDFIELD data was used in a study that demonstrated that the appearance of a high dropout rate in engineering was really the result of a higher-than-typical retention rate and a low replacement rate compared to other disciplines [3]. Research with MIDFIELD showed that women are as likely as men to persist in engineering even when disaggregated by race and sex, and that women follow similar pathways to men if they leave engineering [4]. Several studies have illustrated how student demographics and outcomes vary by engineering discipline, gender, and race/ethnicity [5, 6, 7, 8, 9, 10, 11, 12]. A detailed study of four-year vs six-year graduation rates illuminated a "systematic majority measurement bias" [13]. Contrary to stereotypes of students who change majors, research with MIDFIELD data revealed that students who migrate into

engineering disciplines are successful [14]. In conducting this research with such a rich and detailed dataset, researchers have also established new metrics such as "stickiness" [15] and migration yield [16]. MIDFIELD studies using a proxy for socioeconomic status (SES) revealed that race/ethnicity was no longer as strong a predictor of success when SES was considered [17]. In comparing pipeline, pathways, and ecosystems metaphors for studying student persistence, researchers used MIDFIELD data to illustrate metrics that are informed by each metaphor [16].

Research using MIDFIELD has been recognized with many invitations for invited talks and awards demonstrating the quality of the work and its importance to the engineering education research community. Awards for work with MIDFIELD include the 2013 Women in Engineering Proactive Network (WEPAN) Betty Vetter Award for Research "for exceptional research committed to understanding the intersectionality of race and gender" [18] and five journal best paper awards: the 2011 and 2015 IEEE *Transactions on Education* Theodore E. Batchman Best Paper Awards [7, 8, 19] and the 2008, 2011, and 2019 William Elgin Wickenden Award for the Best Paper in the *Journal of Engineering Education* (JEE) [3, 13, 16, 20].

III. EXPANDING NETWORK OF MIDFIELD RESEARCHERS

The MIDFIELD team aims to educate the broader research community, expand the network of researchers capable of conducting this research, and share innovative research methods in addition to the actual data. We have offered workshops at several conferences as well as the MIDFIELD Institute.

A. Inagural MIDFIELD Institute

The inaugural MIDFIELD Institute brought together about twenty researchers from across the USA for two days in June 2019. More details are available on the MIDFIELD Institute website [21].

Learning objectives for the MIDFIELD Institute included qualitative and computational ones. Qualitatively, by the end of the workshop participants should be able to:

- Describe the data available in MIDFIELD
- Describe how the MIDFIELD data are organized
- Describe key principles of effective data visualization
- Identify deficiencies of common graph types

Computationally, participants should be able use *midfieldr*, an R package specifically designed for use with MIDFIELD, to:

- Calculate and evaluate educational metrics
- Produce a table of data that addresses a research question
- Explore and tell a story from MIDFIELD data

The *midfielddata* practice data set, a stratified sample of data from MIDFIELD, was designed to facilitate accessing and analyzing MIDFIELD data. This practice data set includes de-identified student records for 97,640 undergraduate students from 1987 to 2016. The R software environment was chosen for several reasons including that it is free, open source, available on every major platform, and offers superior tools for contemporary data graphics [22]. The *midfielddata* R data package is available on GitHub [23]. The *midfieldr* package, also available via GitHub, provides tools for accessing and analyzing these data [24]. For example, researchers can select specific programs to study (e.g. electrical engineering or mechanical engineering), group and summarize by race/ethnicity, sex, and program, or compute and graph various persistence metrics.

For participants who wanted more practice with R and could come early, an optional pre-workshop session included tutorials on the basics of the R environment, graphing in R, and data structures in R. Day 1 began with an introduction to MIDFIELD, the facilitators, and the goals of the MIDFIELD Institute. Then participants worked through self-paced tutorials on *midfieldr* and *midfielddata*. After summarizing some key results and data displays from MIDFIELD research, participants explored defining new research question with facilitators providing examples of specific items to consider. Day 1 ended with self-directed practice of defining a problem involving data of interest to participants. Day 2 began with a discussion of effective data visualization followed by an exploration of data to identify and present stories. Then participants worked on their own research question exploring data displays and creating a poster to describe their current work. The Institute ended with an interactive poster session.

B. Evaluation of Inagural MIDFIELD Institute

Participants were invited to complete an evaluation at the conclusion of the Institute. Thirteen participants completed this evaluation. Results from specific questions about the Institute can be found in Figure 1. "Not applicable" was given as an option but no participants chose this for any question. Overall, the participants were pleased with the Institute and rated it highly. There was high agreement about the value of the content and delivery including activities giving sufficient practice and feedback, stimulating learning, and appropriate pace. Participants felt that instructors were helpful and well prepared. Most participants reported that the instituted lived up to their expectations and was a good way to learn about how to use MIDFIELD data. The lowest agreement was found for the content being relevant to job and able to be used. This suggests that we need to probe more deeply on these areas. The disagreement could be because MIDFIELD is not directly a part of their jobs or they do not feel they understood how to use MIDFELD well enough.

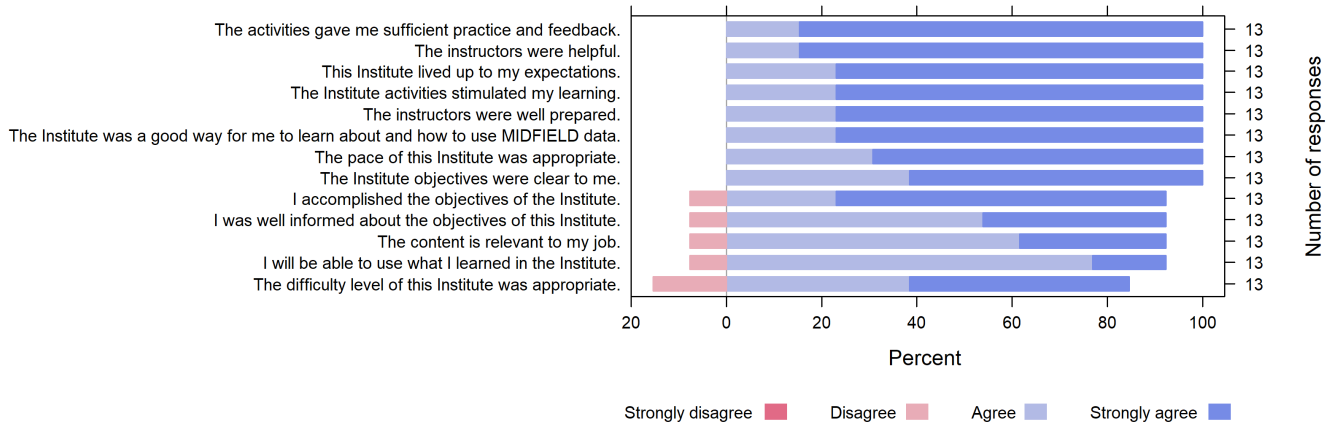**Participant responses to evaluation questions at the MIDFIELD Institute**

FIGURE 1: PARTICPIANT RESPONSES TO EVALUATON QUESTIONS FOR MIDFIELD INSTITUTE (N = 13)

Participants were also asked to choose from a list of possible improvements for the MIDFIELD Institute. They could choose as many items as they wanted. The results are shown in Table I. Note that the participants wanted more content, more time, and more difficulty. Participants also wanted better information before the Institute. Given that this was the first offering, there is definitely room for improvement in preparations and logistics.

TABLE I.   RESPONSES TO "HOW WOULD YOU IMPROVE FUTURE MIDFIELD INSTITUTES" FROM PARTICIPANTS (N = 13)

	Responses	
	#	Percentage
Provide better information before the Institute.	8	62%
Increase the content covered in the Institute.	8	62%
Allot more time for the Institute.	8	62%
Make the Institute more difficult.	7	54%
Clarify the Institute objectives.	5	38%
Make Institute activities more stimulating.	4	31%
Speed up the pace of the sessions.	3	23%
Reduce the content covered in the Institute.	2	15%
Slow down the pace of the sessions.	2	15%
Update the content covered in the Institute.	1	8%
Improve the instructional methods.	1	8%
Make the Institute less difficult.	1	8%
Improve Institute organization.	0	0%
Shorten the time for the Institute.	0	0%

IV. THIS WORKS-IN-PROGRESS SESSION AT FIE2020

Several of the researchers who attended the MIDFIELD Institute are presenting their work at FIE and will be in this works-in-progress (WIP) session. This will provide FIE 2020 attendees with a range of examples of research that is possible with MIDFIELD. Research includes

- H. EbrahimiNejad, C. E. Brawner, and M. W. Ohland, Purdue University and Research Triangle Educational Consultants, "WIP: Demographic and financial trends among southeastern universities in the U.S.A." [25]
- L. Lampe, University of Virginia, "Undergraduate Academic Probation First Semester and Subsequent Academic Performance" [26]
- M. W. Ohland, Purdue University and S. M. Lord, University of San Diego, "The Role of Introductory Course Grades in Engineering Disciplinary Cultures" [27]
- G. Ricco, University of Indianapolis, "MIDFIELD Special Session: A Primer on Novel Methodologies in Longitudinal Analysis of Student Data" [28]

V. SHARING DATA

We are currently working on a partnership with the American Society for Engineering Education (ASEE) that will make ASEE the archivist for MIDFIELD. This will enable ASEE to use MIDFIELD in publishing their DataBytes feature in *Prism* magazine and other publications [29]. This will facilitate feedback mechanisms and a community forum to support researchers using MIDFIELD. We believe that this partnership will help incentivize more institutions to join MIDFIELD to help us reach our goal of approximately 100 institutions containing over 10 million student records. Such a dataset would represent over 50% of the USA engineering undergraduate degrees awarded and increase the diversity of institutions to include public and private institutions, minority-serving institutions, and institutions from a range of research classifications. The scope of MIDFIELD will enable significant improvements in research in higher education.

REFERENCES

[1] M. W. Ohland and R. A. Long, "The Multiple-Institution Database for Investigating Engineering Longitudinal Development: An experiential case study of data sharing and reuse," *Advances in Engineering Education,* vol. 5, no. 2, pp. 1-28, 2016.

[2] M. K. Orr, M. W. Ohland, S. M. Lord, and R. A. Layton, "Comparing the Multiple-Institution Database for Investigating Engineering Longitudinal Development with a National Dataset from the United States," *International Journal of Engineering Education*, vol. 36, no. 4, pp. 1321-1332, 2020.

[3] M. W. Ohland, S. D. Sheppard, G. Lichtenstein, O. Eris, D. Chachra, and R. A. Layton, "Persistence, engagement, and migration in Engineering," *J. Eng. Ed.,* vol. 97, no. 3, pp. 259-278, 2008.

[4] S. M. Lord, M. M. Camacho, R. A. Layton, R. A. Long, M. W. Ohland, and M. H. Wasburn, "Who's persisting in engineering? A comparative analysis of female and male Asian, Black, Hispanic, Native American and White students," *J. Women Minorities Science Eng.,* vol. 15, no. 2, pp.167-190, 2009.

[5] M. Pilotte, M. W. Ohland, S. M. Lord, R. A. Layton, and M. K. Orr, "Student Demographics and Outcomes in Industrial Engineering," *International Journal of Engineering Education*, vol. 33, no. 2A, pp. 506-518, 2017.

[6] M. K. Orr, S. M. Lord, R. A. Layton, and M. W. Ohland, "Student Demographics and Outcomes in Mechanical Engineering in the U.S.," *International Journal of Mechanical Engineering Education*, vol. 42, no. 1, pp. 48-60, 2014.

[7] S. M. Lord, R. A. Layton, and M. W. Ohland, "Trajectories of Electrical Engineering and Computer Engineering Students by Race and Gender," *IEEE Transactions on Education*, vol. 54, no. 4, pp. 610-618, 2011.

[8] S. M. Lord, R. A. Layton, and M. W. Ohland, "Multi-institution Study of Student Demographics and Outcomes in Electrical and Computer Engineering in the U.S.A.," *IEEE Transactions on Education*, vol. 58, no. 3, pp. 141-150, 2015.

[9] C. E. Brawner, S. M. Lord, R. A. Layton, M. W. Ohland, and R. A. Long, "Factors Affecting Women's Persistence in Chemical Engineering," *International Journal of Engineering Education*, vol. 31, no. 6A, pp. 1431-1447, 2015.

[10] S. M. Lord, R. A. Layton, M. W. Ohland, C. E. Brawner, and R. A. Long, "A Multi-institution Study of Student Demographics and Outcomes in Chemical Engineering," *Chemical Engineering Education*, vol. 48, no. 4, pp. 223-230, 2014.

[11] M. K. Orr, N. M. Ramirez, S. M. Lord, R. A. Layton, and M. W. Ohland, "Student choice and persistence in Aerospace Engineering," *Journal of Aerospace Information Systems (JAIS)*, vol. 12, no. 4, pp. 365-373, 2015.

[12] M. W. Ohland, S. M. Lord, and R. A. Layton, "Student Demographics and Outcomes in Civil Engineering in the U.S.," *Journal of Professional Issues in Engineering Education and Practice*, vol. 141, no. 4, pp. 1-7, 2015.

[13] M. W. Ohland, C. E. Brawner, M. M. Camacho, R. A. Layton, R. A. Long, S. M. Lord, and M. H. Wasburn. "Race, gender, and measures of success in Engineering Education," *J. Eng. Educ.,* vol. 100, no. 2, pp. 225-252, 2011.

[14] S. M. Lord, M. W. Ohland, R. A. Layton, and M. M. Camacho, "'All who wander are not lost.' Examining outcomes for migrating engineering students using ecosystem metrics," *Proceedings of the 2018 Frontiers in Education (FIE) Conference*, San Jose, CA, October 2018.

[15] M. W. Ohland, M. K. Orr, S. M. Lord, R. A. Long, and R. A. Layton, "Introducing 'Stickiness' as a Versatile Metric of Engineering Persistence," *Proceedings of the 2012 Frontiers in Education (FIE) Conference*, Seattle, WA, October 2012.

[16] S. M. Lord, M. W. Ohland, R. A. Layton, and M. M. Camacho, "Beyond Pipeline and Pathways: Ecosystem Metrics," *Journal of Engineering Education*, vol. 108, no. 1, 2019. DOI: 10.1002/jee.20250

[17] M. W. Ohland, M. K. Orr, V. Lundy-Wagner, C. P. Veenstra, and R. A. Long, "Viewing access and persistence in engineering through a socioeconomic lens," in *Engineering and Social Justice: In the University and Beyond*, C. Baillie, A. L. Pawley, and D. Riley, Eds., Purdue University Press, 2012 , 157-182.

[18] 2013 Betty Vetter Award http://www.wepan.org/page/WEPANAwardsForm#Betty%20Vetter%20Award%20for%20Research

[19] http://ieee-edusociety.org/awards/ieee-transactions-education-theodore-e-batchman-best-paper-award

[20] https://www.asee.org/member-resources/awards/full-list-of-awards/national-awards/special

[21] https://midfieldr.github.io/workshops/midfield-institute.html

[22] R Core Team, "R: A language and environment for statistical computing," R Foundation for Statistical Computing, Vienna, Austria, 2018. [Online]. Available: https://www.R-project.org/

[23] R. Layton, R. Long and M.Ohland, "midfielddata: Student Record Data for 98,000 Undergraduates," 2018. R package version 0.1.0. https://github.com/MIDFIELDR/midfielddata

[24] R. Layton, R. Long and M. Ohland, "midfieldr: Tools for Student Records Research," 2018. R package version 0.1.0.9002. https://github.com/MIDFIELDR/midfieldr

[25] H. EbrahimiNejad, C. E. Brawner, and M. W. Ohland, "WIP: Demographic and financial trends among southeastern universities in the U.S.A." *Proceedings of the 2020 Frontiers in Education (FIE) Conference*, Uppsala, Sweden, October 2020.

[26] L. Lampe, "Undergraduate Academic Probation First Semester and Subsequent Academic Performance", *Proceedings of the 2020 Frontiers in Education (FIE) Conference*, Uppsala, Sweden, October 2020.

[27] M. W. Ohland and S. M. Lord, "The Role of Introductory Course Grades in Engineering Disciplinary Cultures," *Proceedings of the 2020 Frontiers in Education (FIE) Conference*, Uppsala, Sweden, October 2020.

[28] G. Ricco, "MIDFIELD Special Session: A Primer on Novel Methodologies in Longitudinal Analysis of Student Data," *Proceedings of the 2020 Frontiers in Education (FIE) Conference*, Uppsala, Sweden, October 2020.

[29] ASEE Databyes, available at https://ira.asee.org/recent-news/