Advances in Engineering Education

SPRING 2016

The Multiple-Institution Database for Investigating Engineering Longitudinal Development: an Experiential Case Study of Data Sharing and Reuse

MATTHEW W. OHLAND

AND

RUSSELL A. LONG
Purdue University
West Lafayette, IN

ABSTRACT

Sharing longitudinal student record data and merging data from different sources is critical to addressing important questions being asked of higher education. The Multiple-Institution Database for Investigating Engineering Longitudinal Development (MIDFIELD) is a multi-institution, longitudinal, student record level dataset that is used to answer many research questions about how students maneuver through required engineering curriculum and what courses or policies stand in their way toward graduation. The process of designing, compiling, maintaining, protecting, and sharing a large dataset like MIDFIELD provides valuable insight for others.

Key words: Data sharing, MIDFIELD, student records.

INTRODUCTION

Retention, measured in various ways, has been the dominant mode of studying student success in engineering education and higher education in general. Data available from the Integrated Postsecondary Education Data System (IPEDS, 2015) and discipline-specific sources such as the American Society for Engineering Education (ASEE, 2015), the Engineering Workforce Commission (EWC, 2015), and the National Science Foundation (NSF) Science and Engineering Indicators (NSF SEI, 2015) do not facilitate longitudinal studies – nor were they designed to do so. Student-level longitudinal data allows calculation of six-year graduation rates and many other outcomes. Since such data are rarely available, various alternative measures have been used. Short-term measures such as one-year or two-year persistence fail to capture the important outcome of graduation, while cross-sectional approaches make the risky assumption that each cohort is the same as the next (Cosentino de Cohen & Deterding, 2009) and cannot account for migration among majors. The extensive use of data from the National Educational Longitudinal Study (NELS, 1988, 2002; USDOE NCES, 2000; Rowan, Chiang & Miller, 1997; Dávila & Mora, 2007), which tracks high school cohorts, demonstrates the value of longitudinal data in studying educational outcomes. NELS has been used to study varying topics: gender differences in mathematics achievement (Fan, Chen & Matsumoto, 1997); religious involvement, social capital, and adolescents' academic progress (Muller & Ellison, 2001); adolescent cigarette smoking in U.S. racial/ethnic subgroups (Johnson & Hoffmann, 2000); and the social networks and resources of African–American eighth graders (Smith-Maddox, 1999).

Researchers in the field of engineering education understood the need for and created a longitudinal student database that can be used to study how engineering students move through the curriculum and to create national benchmarks (Carson, 1997; Ohland & Anderson, 1999). The Southeastern University and College Coalition for Engineering Education (SUCCEED) Longitudinal Database (LDB) was created in 1996. SUCCEED was one of eight coalitions developed by the National Science Foundation (NSF) through the Engineering Education Coalition (EEC) program. Out of the success of SUCCEED came the Multiple-Institution Database for Investigating Engineering Longitudinal Development (MIDFIELD, 2015).

MIDFIELD provides longitudinal data for 1,014,887 undergraduate students since fall 1987 – 210,725 of those students ever declared engineering as a major. MIDFIELD comprises whole population data of degree-seeking students at the 11 partner institutions—including students of all disciplines, transfer students, part-time students, and students who first enroll at any time of year. 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 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. 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 HBCUs. The graduation percentage of Hispanics (regardless of gender) 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 nation are awarded to Hispanics. Hispanic students are particularly concentrated at two institutions in the database, Georgia Tech and the University of Florida. Together they account for 65 percent of the Hispanic population in our database. All other racial/ethnic populations are representative of a national sample (Yoder, 2013).

The long-term plan for MIDFIELD has always been to expand to include all institutions in the United States that offer undergraduate programs in engineering. MIDFIELD is growing and has been funded by the National Science Foundation (Ohland, et al., 2016) to initially 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. public institutions and approximately two-thirds of the U. S. undergraduate engineering enrollment in any given year. 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 missions.

The expansion of MIDFIELD brings with it reflection on the establishment, design, maintenance, protection, and sharing of the database. These reflections should provide valuable insight for other researchers.

THE SUCCEED COALITION AND THE FOUNDATION OF A DATA PARTNERSHIP

The Southeastern University and College Coalition for Engineering Education (SUCCEED) Longitudinal Database (LDB) was created in 1996. SUCCEED was one of eight coalitions developed by the National Science Foundation (NSF) through the Engineering Education Coalition (EEC) program.

> Through the EEC program groups of universities and colleges of differing characters formed Coalitions in order to become change agents amidst the engineering education community. Goals for systemic reform included increased retention of students, especially underrepresented groups such as white women and underrepresented minorities, improved introductory experiences in engineering, active experiential learning experiences such as artifact dissection, and multidisciplinary capstone design experiences. (EEC, 2005)

The LDB contained undergraduate student data from the nine SUCCEED partners, all southeastern, public universities: Clemson University, Florida A & M University, Florida State University, Georgia Institute of Technology, North Carolina Agricultural & Technical State University, North Carolina State University, University of Florida, University of North Carolina Charlotte, and Virginia Polytechnic Institute and State University. The LDB contained demographic, attendance, and graduation data files for all undergraduate students from all nine universities.

When the LDB was first created, most institutions used students' social security numbers to synchronize records. The software used to manipulate and store the data was SAS. Data files were stored as SAS7BDAT files – a binary database storage file. SAS software was chosen because it has the flexibility and power to easily accomplish all the main project tasks, including input of raw text files, manipulation of data elements, working with datasets containing millions of records, and production of standard reports and statistical analyses.

The SUCCEED Coalition partnership forged a relationship among its member institutions, but the initial impetus to create a longitudinal database with contributions from each of the partners was a request from National Science Foundation officials to demonstrate the benefits of SUCCEED during the Coalition's 4th year review. The hope of funding to continue the Coalition's work provided a powerful incentive to create a data partnership, garnering support from powerful allies. Letters from each institution supporting the creation of the SUCCEED Longitudinal Database were signed by university-level administrators of each institution, the Engineering Dean at each institution, and the chair of every engineering department at each of the partners. The letters of support are available under each institution's page under the "MIDFIELD Institutions" tab on the MIDFIELD homepage (MIDFIELD, 2015). In a climate of data sharing, transparency seems an important principle, including making public the stakeholders of the data sharing process.

THE TRANSITION TO MIDFIELD

In 2002, negotiations with SUCCEED member institutions resulted in a partnership to extend the LDB. "Studies using the Multiple-Institution Database for Investigating Engineering Longitudinal Development (MIDFIELD)" began in June 2004 with NSF support (NSF REC-ROLE / STEP 0337629, Matthew Ohland PI, $1,470,391, June 1, 2004 to April 30, 2010, including a transfer of institution and a no-cost extension). This project replaced Social Security Numbers with internal identifiers and compiled data from 1987-2005 from all institutions. The SUCCEED partners had promised to continue to provide data to the LDB until three years after the end of the SUCCEED Coalition in August 2003 (MIDFIELD, 2015). The ongoing unfunded burden of supplying data to the LDB made joining MIDFIELD attractive to the SUCCEED partners, since the partners would receive funding to compile the new MIDFIELD dataset (starting anew to avoid any connection to the data linked by Social Security Numbers), and provide an institutional contact to who would assist MIDFIELD staff during data validation. MIDFIELD added data fields identified as useful, but not collected in the LDB, particularly a new course table. MIDFIELD collected the following data tables (* after the variable name indicates that the variable was added during MIDFIELD data collection):

- **Demographic:** reporting institution, term identifier, person identifier number, admissions major code expressed as an NCES Classification of Instructional Programs (CIP) code (NCES, 1990, 2000, 2010), type of student at time of application, racial/ethnic group, gender, matriculation term and year, matriculation major code, high school grade point average (GPA), SAT scores, and ACT scores, fee classification/residency, last institution code, high school code, cumulative hours accepted (transfer), birthdate (day, month, year)*, country of citizenship*, United States visa type*, State of residency at time of entry*, home zip code*, high school rank*, high school size*, veteran status*; (one record per student).

- **Term:** reporting institution, term identifier, person identifier number and data that change each term including student classification level, institutional classification level of student, cooperative education flag, termination code, term credit hours for GPA, cumulative hours for GPA, total cumulative grade points, term grade points earned, current term course load, major during term expressed as a CIP code*, institution code (not CIP) that describes the term major*, on or off campus housing*; (one record per student per enrolled term).
- **Graduation:** records for each bachelor's degree awarded to each student in the database including reporting institution, term identifier, person identifier number, degree level, degree major code expressed as a CIP code, graduation major expressed as a name*; (zero or more records per student).
- **Course***: reporting institution, term identifier, person identifier number, course name, course alpha identifier, course number, course name, course section identifier, course grade, course credits, course method, academic rank of person teaching the course, pass/fail, credit by advanced placement; (one record per student per enrolled course).

Data were validated by creating frequency tables of institutional data, comparing those tables to institutional fact books, and discussing any data anomalies with the assigned institutional representative. The new data fields made it possible to explore a variety of new research questions. Adding age provided valuable information to understand nontraditional students (Bushey-McNeil, Ohland, & Long, 2014; McNeil, Ohland, & Long, 2014). Adding home zip code at matriculation allowed the association of a variety of census data that might be used in socioeconomic modeling (Ohland, Orr, Lundy-Wagner, Veenstra, & Long, 2012; Lundy-Wagner, Veenstra, Orr, Ramirez, Ohland, & Long, 2014). Adding major each term allowed tracking complete student pathways – providing valuable information about both the choices students made and what events might have triggered those choices (Ohland, Brawner, Camacho, Long, Lord, & Wasburn, 2011; Lord, Layton, & Ohland, 2011; Orr, Lord, Layton, & Ohland, 2014; Lord, Layton, Ohland, Brawner, & Long, 2014; Lord, Layton, & Ohland, 2015; Ohland, Lord, & Layton, 2015; Orr, Lord, Layton, Ramirez, & Ohland, 2015). The addition of course data enabled a wide range of research questions (e.g., Ricco, Salzman, Long, & Ohland, 2012).

Purdue University and the University of Colorado joined MIDFIELD in 2010 – bringing the total number of MIDFIELD institutions to eleven. The choice of MIDFIELD institutions was originally targeted to SUCCEED institutions. Adding institutions to MIDFIELD was hindered by lack of funding. The addition of Purdue University and the University of Colorado was by convenience and interest. The MIDFIELD project moved to Purdue University in 2006. Once the project staff had established themselves on campus, negotiations began to add Purdue data to MIDFIELD. The University of Colorado joined MIDFIELD due to administrative interest in potential policy impact and to expand opportunities for doctoral student and faculty research.

CATALOGS AND INSTITUTIONAL POLICIES

Institutions provide MIDFIELD with undergraduate course catalogs and bulletins for each year of student data transmitted. From these catalogs MIDFIELD staff assembled a course database used to track students through the prescribed curriculum. To date, only Science, Technology, Engineering, and Mathematics (STEM) disciplines have been mapped – but there is hope that other researchers will build on this work to eventually include all disciplines.

The catalogs collected also provide historical policy context for the student records data. A study comparing academic policies related to academic good standing, probation, suspension, and expulsion at nine MIDFIELD institutions over 17 years provided a benchmark to which others can compare (Brawner, Frillman & Ohland, 2010). The print and on-line versions of the undergraduate catalogs from 1988-2005 for each of the institutions were examined regarding those policies. Each school required a 2.0 cumulative grade point average (CGPA) for graduation, but students earlier in their careers remained in good standing with lower CGPAs that vary by institution. Students not in good standing might have been put on probation while remaining in school and given a chance to improve their grades. Failing that, they might have been suspended with various paths to return. After one or two suspensions, students were expelled, although six institutions had policies allowing them to return after time away. Grade forgiveness policies were also examined. Over time those institutions with lower standards for remaining in good standing had raised them. This points to the importance of chronicling institutional policy and how policy affects students' academic progress. Without having an understanding of these effects, the reliability and validity of the data might be called into question.

The Protection Of Human Subjects

The focus of discussions regarding human subjects is always two-sided: while many researchers focus on the management of risk, demonstrating how the research will have benefits—in the broadest sense—is also important. That impact of the research can be on the participants themselves, educators, advisors, policy makers, funding agencies, and others. Some MIDFIELD work makes the claim that the general public will benefit from a greater understanding of engineering. A press release announced that engineering had less attrition than other groups of disciplines (Research Findings, 2010) in the hope that parents wouldn't discourage their children from pursuing engineering because it was "a weed-out major." While the press release probably didn't change widespread public perceptions, it did get picked up by 62 media outlets including online stories by the Chicago Tribune, InformationWeek, and US News & World Report.

There are always FERPA concerns with a data set that collects the type of data contained in MIDFIELD. Will collecting this data affect students' "rights and welfare"? The justification for MIDFIELD to collect and analyze this type of data can be found in 20 USC §1232g(b)(1)(F):

> *(b) Release of education records; parental consent requirement; exceptions; compliance with judicial orders and subpoenas; audit and evaluation of Federally-supported education programs; recordkeeping.*
>
> *(1) No funds shall be made available under any applicable program to any educational agency or institution which has a policy or practice of permitting the release of educational records (or personally identifiable information contained therein other than directory information, as defined in paragraph (5) of subsection (a)) of students without the written consent of their parents to any individual, agency, or organization, other than to the following—*
>
> *(F) **organizations conducting studies for, or on behalf of, educational agencies or institutions for the purpose** of developing, validating, or administering predictive tests, administering student aid programs, and **improving instruction**, if such studies are conducted in such a manner as will not permit the personal identification of students and their parents by persons other than representatives of such organizations and such information will be destroyed when no longer needed for the purpose for which it is conducted;*

MIDFIELD contracts with the member institutions through primary agreements or Memoranda of Understanding. Specific limitations and responsibilities are described in those documents. These MOU protect the confidentiality of both students and institutions. Reports aggregated by student and institution will be made available. Where data are disaggregated by institution, the identity of the institutions is masked. MIDFIELD strictly abides by individual State statutes in regard to student data security and confidentiality. Institutions do not provide data for students who have notified the college registrar that they do not want disclosure of directory information without prior written consent.

Data Security

The computers on which MIDFIELD data resides are not networked or connected to the internet. Member institutions transmit data to the MIDFIELD data steward via password-protected, encrypted files. Physical files are stored in a locked filing cabinet in a secure office. Only the MIDFIELD data steward and project director have access to these files. Student identifiers are created especially for MIDFIELD – they are not Social Security Numbers or student IDs. MIDFIELD data has been cleaned and verified.

Student Confidentiality

Data security is only the beginning of the protection of student data. Ironically, the very fact that MIDFIELD has student records for over one million students makes it easier to protect the confidentiality of individual students—their identity is protected primarily by reporting only aggregated results. While research using MIDFIELD conforms to standard cell-size limitations in its research designs (NCES, 2002) the large population in MIDFIELD frequently permits the adoption of stricter minimum cell sizes that both protect students and give greater confidence in the results. MIDFIELD researchers are generally bound by a minimum cell size of 10. Furthermore, MIDFIELD researchers avoid reporting too much information about groups of students. For example, when discussing outcomes of a population that is disaggregated by race/ethnicity, gender, and discipline, aggregating those students across multiple institutions researchers can provide protection for both students and institutions. Researchers are discouraged from using MIDFIELD data to predict the behavior or outcomes of an individual, which results in the ecological fallacy. MIDFIELD cannot predict what a student will do. MIDFIELD is best used to show what large numbers of students have done.

Institutional Context

To avoid harming the institutional partners and MIDFIELD's relationship with them, validation of both the MIDFIELD dataset and any results released publicly are critical. This has resulted in a long learning curve for new researchers (including the authors) in developing both the expertise and the confidence to publish results. The challenges extend well beyond knowledge of data management and statistical procedures in a general sense. The primary challenge lies in MIDFIELD-specific issues of merging data from institutions that have different data handling practices, different data schema, different academic policies, and different institutional histories. Once the data are in the MIDFIELD common format, many differences in data-handling practices have been smoothed over, yet some remain. Students who are planning to pursue engineering but have yet not selected a specific discipline are tracked ways that vary by institution, when the student expresses interest in engineering, and whether they meet engineering's admission standards. How this is sorted out can affect reported matriculation patterns and retention rates. Institutions track participation in cooperative education in different ways, so there is a difference between whether a student is on co-op in a particular term (in the term table) and whether a student has ever participated in the co-op program (a logistic variable in the demographic table). Each institution has policies regarding academic probation, suspension, expulsion, and readmission, but the criteria defining each of those differ as do each institution's related supports and consequences for students. Some of the biggest challenges are those of institutional context, because those have been learned over a long period of partnership with the institutions and through more intensive interviews of knowledgeable personnel at institutions joining the partnership more recently. Researchers might be surprised that no students graduated from Georgia Tech in Summer 1996, until they are reminded that Atlanta hosted the Olympic and Paralympic Games and that the Olympic Village that housed visiting athletes occupied most of the Georgia Tech campus. The learning curve involved in using MIDFIELD safely and effectively is an ongoing challenge for the project, particularly as we seek to share MIDFIELD data with an ever-larger community of researchers, including those with whom we might not have direct contact.

Institutional Confidentiality

The MIDFIELD partner institutions must be protected. The most dramatic expression of this sentiment came from an administrator at an institution considering joining MIDFIELD: "We want to make sure our data isn't weaponized." Participants in interviews and focus groups identified a variety of negative outcomes:

- Judging an institution by metrics that do not measure what the institution values.
- Comparing an institution to others using a metric that intends to measure what the institution values, but where the metric is defined in a way that favors other institutions.
- Releasing information that might provide competing institutions an advantage. This concern was most acute among schools competing for funding within a state system.
- Focusing on student outcomes without regard for their initial preparation.

MIDFIELD avoids linking findings to specific institutions. Many of the methodological approaches used to conceal institutional identity are mundane—reporting data using percentages rather than raw data to mask institution size, producing multiple graphs in the same publication using a different institutional key each time to avoid the cumulative loss of anonymity, aggregating data across institutions, and other approaches. Findings have sometimes been linked to policies, but only where those did not betray institutional confidentiality (because those policies were common to multiple institutions). We strive to explore institutional variability without compromising three important principles (Ohland, Brawner, Camacho, Long, Lord, & Wasburn, 2011):

1. Institutional data are provided to the MIDFIELD project on the condition that researchers protect the identity of the partner institutions and each institution's students.

2. Increasingly specific institutional descriptions discourage readers from considering MIDFIELD research to be generalizable, in spite of other significant evidence that there is much that is common among engineering programs and their interaction with students.
3. While MIDFIELD includes data for very large numbers of students, a relatively small number of institutions are represented, so institutional variation must be treated using a case study approach. Conscientious institution-level analysis would require a large number of diverse institutions.

This last principle can be difficult in cases where it appears that something is specific to an institution, a particular type of institution (Historically Black Colleges and Universities), or institutions that share a particular policy. When researchers (including the authors) begin to speculate along those lines, others on the research team are expected to recall one of the team's catchphrases: "If the institution is the unit of analysis, we only have a sample size of 11." As we compare outcomes of first-year engineering programs with those of institutions where students matriculate directly to a discipline, this is a persistent limitation in our work.

Benefits To Institutional Partners

Institutional partners consider the research results published from MIDFIELD to be an important return on their investment in the partnership, and some of our institutional partners follow, question, and act on our findings. The best form of reciprocity in this case comes in the form of direct benefits available only to the partner institutions. MIDFIELD data are more accessible to the partner institutions, either through small subcontracts to compensate the data steward in generating datasets and mentoring researchers using MIDFIELD or by seeking letters of support needed for a project as part of the partnership. As indicated previously, the learning curve for using MIDFIELD data remains a challenge in this area, even when data dictionaries are available to experienced researchers. Occasionally, special requests have been honored for peer comparisons where the confidentiality of the other institutions can be protected. A series of special reports were released in 2011 that was customized for each partner institution, contained peer comparisons, and even made institution-specific recommendations. Similarly, when MIDFIELD results are published in which institutional comparisons are made anonymously, we transmit to each institution a key to its identity in each data display.

Benefits to the engineering education community

A considerable amount of research has been conducted using MIDFIELD, resulting in more than 20 publications in journals and more than 80 in conference proceedings, more than 30 other conference presentations, a book chapter (Ohland, Orr, Lundy-Wagner, Veenstra, & Long, 2012), and a book (Camacho & Lord, 2013). The quality of research using MIDFIELD has been recognized with the best paper award in the Journal of Engineering Education in 2008 and 2011 and the best paper in the IEEE Transactions on Education in 2011 (Ohland, et. al, 2008; Ohland, et. al, 2011; Lord, Layton, & Ohland, 2011). MIDFIELD colleagues have also received best paper awards at two national conferences (Zhang, et. al., 2003; Ohland, Zhang, Thorndyke, & Anderson, 2004). MIDFIELD results have been disseminated through participation in panels (Batchman, et. al., 2005; Long & Ohland, 2011; Brawner, et. al., 2011), an invited workshop at an NSF grantees meeting (Ohland, 2009), four keynote addresses (Ohland, 2005; Lasser & Ohland, 2003; Ohland, 2012; Lord, 2010), more than 20 invited talks, and various media outlets (Basken, 2009). MIDFIELD researchers have been particularly successful in studying the impact of race, socioeconomic status, and gender on success in engineering education and were recognized by the Women in Engineering ProActive Network for exceptional research committed to understanding the intersectionality of race and gender.

In addition to these recognitions, MIDFIELD research benefits the engineering education community by providing results that change conversations—for evidence that engineering is not necessarily a weed-out major and for evidence that there is not necessarily a gender gap in persistence during the college years—these findings provide valuable baseline information. Two policy changes that have stemmed from MIDFIELD research follow.

- MIDFIELD research found that one institution had students who were being retained for a long period, but never graduated. Students were being allowed to progress with CGPAs below the requirement for graduation. Institutional administrators changed the probation and progression requirements – a change that was good for both the institution and for students who were struggling in a major in which they would never graduate.
- MIDFIELD research found that the switch into engineering for students who begin college in a non-engineering major is difficult because required math and science courses in engineering courses are often engineering specific – engineering calculus was different than business calculus. This difference required students switching into engineering to have to retake calculus. One MIDFIELD institution changed policy to begin offering a general calculus course, making switching easier.

METHODOLOGICAL ISSUES AND DEVELOPMENTS IN MIDFIELD RESEARCH

MIDFIELD researchers have faced criticism from some reviewers who expect (or demand) that we publish the results of statistical tests. Various specialized statistical procedures are appropriate and have been used in the study of MIDFIELD, including systematic stepwise regression (Zhang, Anderson, Ohland, Carter, & Thorndyke, 2004), multi-level modeling (Padilla, Zhang, Anderson

& Ohland, 2005; Ricco, Salzman, Long, & Ohland, 2012), and survival analysis (Min, Zhang, Long, Anderson, & Ohland, 2011). Nevertheless, much of the research using MIDFIELD leads to comparing one or more outcomes in various populations. Such questions typically use inferential statistics that are designed to infer whether the differences observed in a sample are likely to be present in the population. Since MIDFIELD includes whole population data for the partner institutions, there is no need to infer anything about the population behavior—it is known. The challenge to this is the extent to which the institutions and students included in MIDFIELD are representative of other institutions in the United States. As stated earlier, compared to other U.S. institutions offering engineering, the MIDFIELD institutions have higher engineering enrollments and a higher fraction of engineering students on campus. Black students are overrepresented, Hispanic students are underrepresented, only public institutions are represented, and the current MIDFIELD partners are predominately in the Southeast. So it is possible that analyses of MIDFIELD may not be generalizable to national data or even data from institutions in other states/regions of the United States—they may only be representative of the MIDFIELD institutions. No national data are available to test this assumption. To the extent that the results from MIDFIELD are representative of other engineering institutions, they are most likely representative of other large, public institutions.

Another challenge researchers have faced is connecting MIDFIELD work to previous research. In some cases, literature from outside engineering education provides a valuable backdrop for findings. In other cases, researchers are able to place their work in the context of findings from other researchers, even if the dataset used in those earlier studies lacked the longitudinal, multi-institutional, and large-population characteristics of MIDFIELD. In some cases, however, MIDFIELD researchers face a challenge in finding suitable comparators, and reviewers charge insularity—of showing a lack of respect for other published work. The more persistent challenge faced is that institutional differences in policy, calendar, and curricular structure create methodological issues in comparing institutions.

Diversity of matriculation models.

In studying MIDFIELD, it is clear that the diversity of matriculation models affects data architecture, study design, and data interpretation. If we compute the retention rate in a particular engineering discipline at an institution with a first-year engineering program, the attrition of students who leave engineering before choosing a discipline will not be counted, resulting in systematically higher retention rates. Matriculation models include first-year engineering programs (where students cannot select a specific major until they complete the first year), selecting a specific engineering major when applying, and multiple models in between. To pool data from institutions with disparate models, researchers must account for these differences. One approach MIDFIELD researchers have used is to impute the number of students in each major at matriculation (Orr, Lord, Layton, & Ohland, 2014;

Lord, Layton, Ohland, Brawner, & Long, 2014; Lord, Layton, & Ohland, 2015; Ohland, Lord, & Layton, 2015; Orr, Lord, Layton, Ramirez, & Ohland, 2015).

Diversity of term structure and measures of student progression

IPEDS identified the diversity of term configurations in higher education as a challenge for the timing of reporting (Cunningham & Milam, 2005). The issue of greater concern in this case is the impact that term structure diversity has on methodology. In the study of MIDFIELD, it has become clear that there are decisions to be made in how student progression is measured, and that these decisions have consequences (Ohland, Brawner, Camacho, Long, Lord, & Wasburn, 2011). One set of combinations in measuring progress are semesters/quarters/trimesters with/without counting summer terms and counting/not counting terms in which the student is/is not enrolled. Focusing on student enrollment rather than chronological time remove the sensitivity of the progression metric to students who "stop out," a practice that is more common among minority students (Love, 1993). It is also possible to measure simply "terms" (regardless of their length), since students have choices to make (regarding their major and what classes they will take) at the end of each term. In that sense, students who attend an institution on the quarter system are faced with (or forced to face) the decision to stay in or leave engineering more times in their academic career than a student on the semester system. Chronological time (years) is important in the context of graduation rates, which typically allow students 150 percent of nominal time to completion—a six-year graduation rate in the case of a four-year undergraduate degree program. Chronological time is also relevant to time-to-degree metrics, financial aid and scholarship eligibility, and lost wages. Credit hours completed are a useful measure of academic progress, but can be misleading where students have accumulated credits that do not count toward graduation. For example, the achievement of certain milestones can be used as a progression metric. MIDFIELD researchers have used completion of a first-year engineering program and entry into or completion of a course or sequence as progression metrics.

Creating new metrics

To address the challenges of reporting results from diverse institutions, MIDFIELD researchers have already proposed some new metrics:

- The "stickiness" of a major is how likely students are to "stick" to that major once they choose it—regardless of what other majors they have had, what other institutions they have attended, or how long they have been in college when they first enroll in that major. The stickiness of a major is the number of students who graduate in that major divided by the number of students who have ever been enrolled. This metric requires a single assumption: that selecting a major indicates intent to graduate in that major (Ohland, Orr, Layton, Lord & Long, 2012).

- Peer Economic Status (PES) is a measure of the average economic status of a student's high school, is a significant predictor of college persistence (Orr, Ramirez, & Ohland, 2011; Ohland, Orr, Lundy-Wagner, Veenstra, & Long, 2012; Orr, Ramirez, & Ohland, 2012). The PES variable is coded so that a higher value corresponds to a better peer economic status, which is a number between 0-100.
- MIDFIELD researchers hope to develop a "percent of degree program completed" metric that would classify student progression consistently at any point in time, regardless of mode of entry, and regardless of speed of completion. This will make it possible to compare full-time and part-time students, first-time-in-college and transfer students, and students who switch majors. When students switch majors, this metric would need to be recalculated on the basis of the new major. The challenge of this metric is the time it takes to map all the courses in all the curricula in all the years for all the majors at all the institutions in the dataset. This process is very intensive and unlikely to be scalable to a nationwide dataset.

EXTENDING THE REACH OF MIDFIELD USING OTHER DATASETS

As is the case with any dataset, MIDFIELD has its limitations. The most notable is that while MIDFIELD findings may provide compelling evidence of ***what*** students do, knowing ***why*** they do it is typically out of reach. Some MIDFIELD research has included the collection and analysis of qualitative data (Mobley, Brawner, & Ohland, 2009; Brawner, Camacho, Lord, Long, & Ohland, 2012). While partnerships that create a dialog between MIDFIELD and qualitative data sources are valuable, they are difficult to scale and share. Combining data from other large-scale quantitative datasets can also result in a richer dataset in multiple ways—by adding critical context to the dataset, and by providing data on important outcomes that are not represented in MIDFIELD.

Adding context

A variety of datasets have provided context to MIDFIELD resulting in richer interpretations and, in some cases, providing critical information for sense-making.

- ***Peer Economic Status*** - Using high-school codes and home zip codes at matriculation collected as part of MIDFIELD, researchers created socioeconomic variables that could be used in MIDFIELD models by collecting free lunch data (Ralston, Newman, Clauson, Guthrie, & Buzby, 2008) from the National Center for Education Statistics Common Core Data (NCES CCD, 2010). A table to convert high school codes from one format to another was needed, but was not available publicly. MIDFIELD researchers located, secured a copy of the table, and improved the resource, making it more accurate and more complete. A variety of studies have been published focusing on this variable (Ohland, Orr, Lundy-Wagner, Veenstra, & Long, 2012; Lundy-Wagner, Veenstra, Orr, Ramirez, Ohland, & Long, 2014; Orr, Ramirez, & Ohland, 2011; Orr, Ramirez, & Ohland, 2012), and the variable is routinely used in other models.

- ***Understanding the impact of merit-based scholarships*** - Researchers used institutional financial data from publicly available data from the Integrated Postsecondary Educational Data System (Chen, Ohland, & Long, 2013) to understand the impact of merit-based scholarships (Chen & Ohland, 2012)
- ***Developing a taxonomy of matriculation*** - Researchers also obtained publicly available data from ABET (www.abet.org), the American Society for Engineering Education Profiles (profiles.asee.org), and institutional websites to develop a Taxonomy of Matriculation practices as a way to develop a richer understanding of how students develop as engineers. (Chen, Ohland, Long, Brawner, & Orr, 2013)

Documenting new outcomes

While MIDFIELD mostly contains the data that would be included on a student's academic transcript, there are other important outcomes to consider. A particularly problematic issue is that while MIDFIELD contains students' course grades, those are not necessarily an objective test of student learning. Due to different grading practices, the meaning of grades will vary even across sections of the same course (Ricco, Salzman, Long, & Ohland, 2012). In classes where faculty use criterion-referenced grading, student grades may approach an interval scale. The use of norm-referenced grading reduces the resolution to an ordinal scale. In the broader context of comparing grades across courses, disciplines, and institutions, it is dubious that even an ordinal scale is possible. Other important outcomes are more easily inferred through partnerships with other data providers.

When Ohland was President of Tau Beta Pi, he had regular contact with the Director of Professional Services at the National Council of Examiners of Engineering and Surveying (NCEES), which administers the Fundamentals of Engineering (FE) examination. While there is debate regarding the usefulness of FE scores, the FE is the only objective test of students' engineering knowledge. In 2004, it was realized that a MIDFIELD-NCEES partnership had the potential to provide valuable new outcome data related to MIDFIELD and an important outreach effort for NCEES helping to establishing the use of FE scores in research. NCEES issued a letter to MIDFIELD indicating the documentation that would be required to proceed (NCEES, 2004). The partnership was missing only one thing – someone who would make that project a priority—getting institutional permission, collecting and conditioning the data, and establishing connections between the datasets to answer valuable research questions. Some discretionary funding was identified to pay for the data extraction and programming. A dissertation and subsequent publications (Barry & Ohland, 2012; Barry & Ohland, 2009) explored the relationship between approaches to ethics instruction and outcomes on the FE exam, uncovering some interesting disciplinary effects as well.

THE IMPORTANCE OF DEVELOPING A NATIONAL STUDENT UNIT-RECORD DATA SYSTEM

It might seem that the greatest benefit of adding more institutions to MIDFIELD would be to make the database more representative of the larger population of U.S. institutions offering B.S. degrees in engineering and to make the findings of research using the database more generalizable. Of greater interest, however, is creating the conditions to answer research questions that require or would benefit from an institutional unit of analysis or from the use of multilevel models that include an institutional level. Such studies fall into several categories:

- ***Studies of academic policies.*** Academic policies certainly affect the educational environment. Adding institutions to MIDFIELD would allow researchers to establish clearer links between those policies and the educational outcomes of students.
- ***Studies of curricular structure.*** There is much evidence that the way in which students are introduced to engineering is important. Some of this evidence shows the influence of formal first-year engineering programs (Brawner, et al., 2009) and common introduction to engineering courses (Orr, Brawner, & Ohland, 2013). To conduct a robust study of the influence of curricular structure, the database must include not only a larger number of institutions, but institutions representing a greater diversity of curricular models.
- ***Studies that depend on institution-level variables.*** Studies in this group can measure the influence of such variables as institution size, engineering fraction of enrollment, private vs. public control, and variables related to financial need. While these have been studied using other datasets, there is much to learn from studying these in multilevel models that include both institution and student-unit-record data.

As the number of institutions in MIDFIELD grows, it is also likely to be attractive to a larger research community and have a more notable local impact on the institutional partners.

DESIGNING A NATIONAL STUDENT UNIT-RECORD DATA SYSTEM

Based on input derived from interviews and focus groups with engineering administrators, engineering education researchers, registrars, institutional research staff, and data archivists, four design principles have been identified for expanding MIDFIELD into a national unit-record database.

Data should be accessible to a broader community of researchers

Institutional representatives that were interviewed recognized the benefits of allowing the research community to have access to a national student unit-record data system. In addition to accelerating the work of current engineering education researchers, permitting access to a broader research community would attract the research interest of demographers, sociologists, statisticians, and others to research questions of interest to engineering education.

- ***Partner institutions must not be affected negatively by published research results.*** To protect the partner institutions, names of MIDFIELD institutional partners should not be associated publicly with specific statistics or calculations. Tables and figures displaying results should use labels that mask the identities of institutions in the data. Institution names should be used only when data is aggregated across more than one institution, and only then so long it is not possible to deduce the institutions.
- ***Partner institutions should have special access to conduct peer comparisons.*** Institutional representatives were clearly interested in the opportunity to use MIDFIELD data to conduct peer comparisons in greater detail than they have access to with currently available data. At the same time, they were unwilling to allow other institutions to have that level of access to their data without some indication of shared risk and trust. Further, findings from such studies should not have the opportunity to have a negative effect on institutions. The results from such peer comparisons must be used solely for institutional analysis and only information pertaining to the institution itself may be made public.
- ***All institutions should have equal access to benefit from the MIDFIELD partnership.*** To ensure that MIDFIELD does not become a resource that further privileges schools that have the resources to participate, we must find resources for institutions to extract the historical data needed join the MIDFIELD partnership. Yet admission to the partnership is not sufficient to level the playing field. Well-resourced institutions are more likely to have highly skilled researchers who conduct research and publish findings based on MIDFIELD. This benefit cannot be granted to MIDFIELD partners, but a corollary benefit can be assured – that less-resourced institutional partners benefit when other institutions conduct research using MIDFIELD. For this reason, while published research that generates institutional findings must mask institutional identity, institutions must privately be informed of their own identity. Thus researchers at all institutions using MIDFIELD provide an institutional research benefit to all the MIDFIELD partners.
- ***A valuable partnership in data sharing.*** The Interuniversity Consortium for Political and Social Research (ICPSR, 2015) specializes in handling and sharing large datasets. In partnership with ICPSR, the authors have negotiated a complex restricted-use data dissemination agreement that describes a process by which MIDFIELD partner institutions provide institutional data, MIDFIELD

staff convert the institutional data to the MIDFIELD common format and transmit the common format data to ICPSR, and ICPSR archives the data, administers and enforces data use agreements, and provides access to the data to investigators who have executed data use agreements. Two distinct data use agreements implement these requirements: a “restricted data use agreement for research” and a “restricted data use agreement for institutional analysis”. Signatures are being sought from the current partners, and all institutions that join the partnership in the future will be expected to participate in this archive. ICPSR will control the distribution of archived data and will manage risk through restricted-use data dissemination agreements. MIDFIELD staff will continue to add institutions to the archive as agreements are reached with MIDFIELD partners. Derived variables will be added to the common format during updates. MIDFIELD staff will distribute a smaller “dummy” data file with valid variable values for use in workshops and by researchers who want to explore MIDFIELD before contracting with ICPSR to gain access.

A timeline for expansion of institutional partners and research access

The expansion of institutional participation is limited by trust, politics, and other factors. It is unrealistic to expect that MIDFIELD will ever include data from all the U.S. institutions with baccalaureate programs accredited by the Engineering Accreditation Commission of ABET. Research access to the MIDFIELD dataset is limited by concerns for institutional and individual privacy and the liabilities related to those. In spite of these constraints, there are plans to expand both the number of participating institutions and research access to the dataset.

Expansion of institutional partners

Plans are underway to add at least 92 institutions to MIDFIELD by 2021. In addition to the benefits of a larger institutional sample described earlier, these new partners would add diversity by institution size, geographic region, and control (public/private). New funding by the National Science Foundation (NSF Award # 1545667, $4,260,978.00, 03/01/16 to 02/28/2021) will increase the number of partner institutions to 113. New institutional partners will receive funding to provide and update data. As the database reaches this size, joining the MIDFIELD partnership becomes even more attractive. Twenty institutions have signed letters of support and are ready to submit data to MIDFIELD. New institutions will be targeted to reflect variability in geographic region, institution size as determined by the number of engineering graduates per year, and institutional control (public or private). Institutions will also be targeted who excel or fail at graduating under-represented minorities – plans include adding 5 Historically Black Colleges and Universities (HBCUs), 7 Hispanic Serving Institutions (HSIs), 5 institutions with high Native American populations and 7 universities with high Asian/Pacific Islander populations. Including the current MIDFIELD institutions (11 public institutions with 9 in the Southeast, 1 in the Midwest, and 1 in the West), the expanded MIDFIELD will include the following types of institutions:

By region:

Northeast – 13 private, 11 public

Southeast – 7 private, 23 public

Midwest – 6 private, 12 public

Southwest – 2 private, 9 public

West – 6 private, 14 public

By number of engineering graduates:

Fewer than 300 graduates – 20 private, 21 public

301 to 500 graduates – 10 private, 14 public

501 to 1,000 graduates – 4 private, 18 public

Greater than 1,001 graduates – 16 public

The collection of institutional data will proceed in seven phases – each year adding approximately 20 institutions. Multiple activities occur in each phase as shown in Figure 1.

Succession plan

Along with plans for expansion, a succession plan is being developed for both the MIDFIELD project director and the data steward. As new institutional partners are added to MIDFIELD, some of those new relationships build on existing relationships, but some prospective partners have already approached the MIDFIELD team about joining the project. This is a sign that MIDFIELD researchers have earned the trust of the community through the quality of their work, by the rigorous protection of student and institutional confidentiality, and by respect for the trust that has already been extended by other institutions through the release of student data.

Expansion of research access

Archiving the dataset with ICPSR represents an important long-term solution to expanding research access to MIDFIELD, and institutions that join the MIDFIELD partnership are asked to commit to participating in that archiving process. In the short term, providing access to MIDFIELD creates an unfunded burden for the core research team to

- ensure that the requester can be trusted to treat the data with respect,
- determine that the requester has the skills to work with a dataset like MIDFIELD without publishing spurious findings,

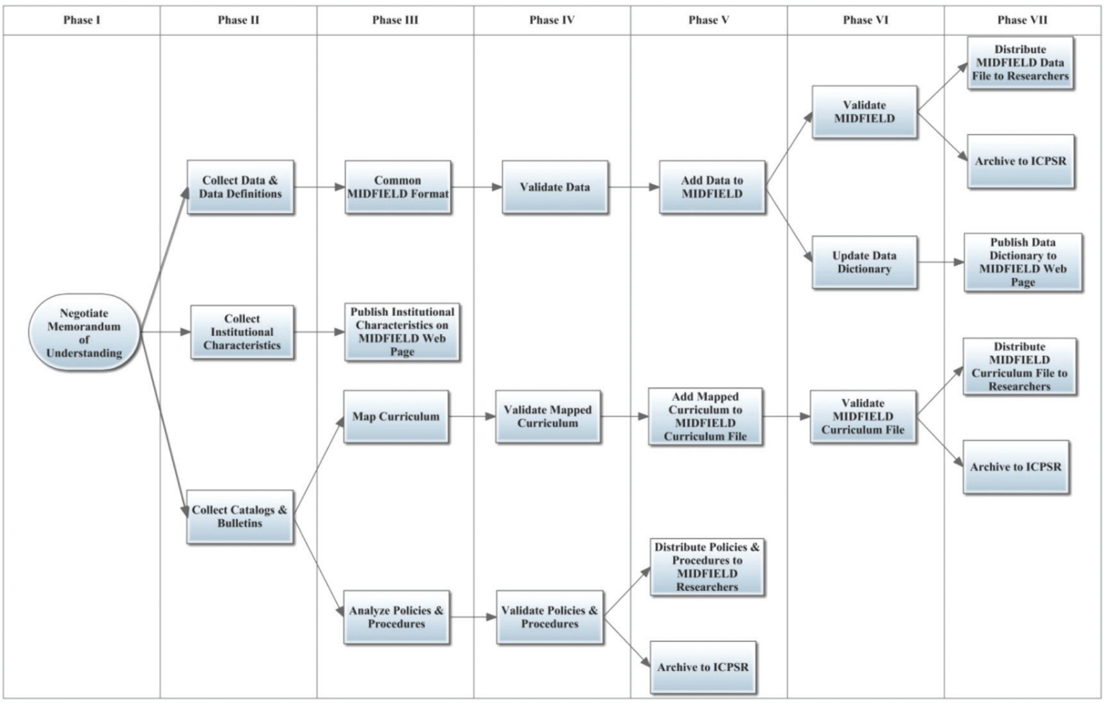

Figure 1. Plan for Institutional Data Collection.

- educate the researcher about the specifics of MIDFIELD, and
- identify and extract the specific data that the requester needs (to avoid releasing more data than is needed, which creates unnecessary risk and is more challenging manipulate).

Various factors mitigate these burdens:

- if there is an incentive for current and future partners to have access,
- if the team has a professional interest in the research and is invited to co-author one or more publications,
- if the researcher can compensate the core research team using institutional or grant funds to offset the burden,
- if a mutually beneficial relationship can be established between a new researcher and others already using MIDFIELD, and
- if a new researcher can provide references or evidence that they will be mentored in their use of the dataset.

Researchers who do gain access to the MIDFIELD data must sign a confidentiality agreement that specifies the terms of use of MIDFIELD data.

CONCLUSIONS AND RECOMMENDATIONS

For eleven institutions, the MIDFIELD team has begun to achieve what NCES concluded could and should be done to form a national student unit data system (Cunningham & Milam, 2005). The clearest messages in the creation, study, and expansion of MIDFIELD relate to the efforts that project personnel have made to build strong relationships of trust and to the measures by which they ensure reciprocity.

Some critical methods to build trust identified in this work are to validate cautiously and publish respectfully. Just as a data dictionary can guide data validation, a clear set of policies for how data are handled in publications establishes a basis for treating students, institutions, and the data with respect. Reciprocity is chiefly established by providing benefits to partner institutions before and after they submit data to the project, but an important consideration is that the need for reciprocity is diminished as the institutional burden is lessened. By accepting data in an institution's native format for project personnel to convert to the MIDFIELD common format, the burden for an institution to provide data to MIDFIELD is significantly reduced. The most significant way to enhance the benefit to the institutional partners is for each partner institution to engage actively and collaboratively in the project. This is the approach taken by the Consortium for Undergraduate STEM Success (CUSTEMS, 2015).

A wide variety of other data sharing strategies are embodied in this work. By developing new metrics suited to the study of MIDFIELD, the database becomes more accessible to researchers. Various strategies are described for adding context to the data—studying institutional policies and curriculum, engaging in qualitative research to explore quantitative findings, and accessing other datasets. Adopting established data management practices enhances scholarly trustworthiness, improves research access, and lessens the burden of maintaining the dataset.

Having a plan for expanding participation, expanding access, and sustaining the database are all important strategies that promote data sharing. Institutions are more likely to join the MIDFIELD partnership if it benefits a larger group of researchers and if it is has a stable future.

REFERENCES

American Society for Engineering Education (ASEE). (2015). Degrees awarded and enrollment reports. ASEE Engineering Data Management System. ASEE: Washington, DC. Retrieved from http://www.asee.org/datamining/reports/

Barry, B.E., & Ohland, M. W. (2009). Engineering ethics curriculum incorporation methods and results from a nationally administered standardized examination. *American Society for Engineering Education 2009 Annual Conference*.

Barry, B.E., & Ohland, M.W. (2012). Criterion 3.f: How much curriculum content is enough? *Journal of Science and Engineering Ethics* **18**(2), pp. 369-392. Retrieved from http://www.springerlink.com/content/05w556u2k5018784/.

Basken, P. (2009). Female students just as persistent as men in engineering, database shows. *Chronicle of Higher Education*, August 4, 2009. Retrieved from http://chronicle.com/article/Female-Students-Just-as/47933/#lastComment

Batchman, T. (Panel Moderator). Panel authors Espinoza, D., White, J., Eschenbach, E., Cashman, E., Ohland, M., Zhang, G., Padilla, M., Anderson, T., Denton, L., McKinney, D., Krause, S., Baur, R., Birk, J., Jenkins, B., & Pavelich, M.. (2005). Using various methods to holistically assess engineering education, *Proceedings of the ASEE/IEEE Frontiers in Education*, Indianapolis, Indiana, October, 2005.

Brawner, C. E., Camacho Walter, M. M., Lord, S. M., Long, R. A., & Ohland, M. W. (2012). Why women choose industrial engineering, *Journal of Engineering Education*, **101**(2) 288–318.

Brawner, C. E., Frillman, S. A., & Ohland, M. W. (2010). A comparison of nine universities' academic policies from 1988 to 2005. (ERIC: ED508293), February 2010, 42 pages.

Brawner, C.E., Camacho, M.M., Frillman, S., Layton, R.A., Long, R.A., Lord, S.M., Ohland, M.W., & Wasburn, M. (2009). Work in progress: The effect of engineering matriculation status on major selection. *Proceedings IEEE/ASEE Frontiers in Education*.

Brawner, C.E., Camacho, M.M., Layton, R.A., Long, R.A., Lord, S.M., & Ohland, M.W. (2011). MIDFIELD researchers discuss pockets of success for women. Panel discussion, Women in Engineering Division, *American Society for Engineering Education 2011 Annual Conference,* Vancouver, Canada, June 2011.

Bushey-McNeil, J., Ohland, M. W., & Long, R. A. (2014). Nontraditional student access and success in engineering. *American Society of Engineering Education 2014 Annual Conference*, Indianapolis, IN.

Camacho, M. M., & Lord, S. M. (2013). *The Borderlands of Education: Latinas in Engineering* Lanham, MD: Lexington Books.

Carson, L. (1997). SUCCEED Quantitative Evaluation. *The Innovator*, No. 8, Fall 1997, SUCCEED Engineering Education Coalition, University of Florida, Box 116134, Gainesville, FL 32611–6134.

Chen, X., & Ohland, M.W. (2012). The effect of college cost and financial aid on access to engineering. *Proceedings of the 2012 American Society of Engineering Education Annual Conference*, San Antonio, TX, June 2012.

Chen, X., Ohland, M. W., & Long, R. A. (2013). The effects of merit-based scholarships on first-year engineering student characteristics and academic behavior. *Proceedings of the 2013 American Society of Engineering Education Annual Conference*, Atlanta, GA, June 2013.

Chen, X., Ohland, M. W., Long, R. A., Brawner, C. E., & Orr, M. K. (2013). A taxonomy of engineering matriculation practices," *Proceedings of the 2013 American Society of Engineering Education Annual Conference*, Atlanta, GA, June 2013.

Consortium for STEM Success (CUSTEMS). (2015). Retrieved from http://www.custems.org/

Cosentino de Cohen, C., & Deterding, N. (2009). Widening the net: National estimates of gender disparities in engineering. *Journal of Engineering Education*, 98(3), 211–226.

Cunningham, A. F. & Milam, J. (2005). Feasibility of a student unit record system within the Integrated Postsecondary Education Data System, NCES 2005-160. Washington, DC: U.S. Department of Education, National Center for Education Statistics.

Dávila, A. & Mora, M. T. (2007). Civic engagement and high school academic progress: An analysis using NELS data. The Center for information & Research on Civic Learning and Engagement, Circle Working Paper 52, January 2007.

Engineering Education Coaltions (EEC). (2005). Retrieved March 18, 2015 from http://www.foundationcoalition.org/home/foundationcoalition/engineering_coalitions.html

Engineering Workforce Commission (EWC). (2015). Retrieved from www.ewc-online.org

Fan, X., Chen, M., & Matsumoto, A. R. (1997). Gender differences in mathematics achievement: Findings from the National Education Longitudinal Study of 1988. The Journal of Experimental Education, 65(3), 229–242.

Integrated Postsecondary Education Data System (IPEDS). (2015). Retrieved from https://nces.ed.gov/ipeds/

Interuniversity Consortium for Political and Social Research (ICPSR). (2015). Retrieved from https://www.icpsr.umich.edu/icpsrweb/content/membership/about.html

Johnson, R. A. & Hoffmann, J. P. (2000). Adolescent cigarette smoking in U.S. racial/ethnic subgroups: Findins from the National Education Longitudinal Study. Journal of Health and Social Behavior, 41(4), 392–407.

Lasser, S.J.S., & Ohland, M.W. (2003). Beating the national average at Clemson University and Math Excellence Workshop at Clemson University. Presentation in Best Practices in Recruitment and Retention session, *National Conference on Best Practices in Black Student Achievement*, Clemson, SC, January 27, 2003.

Long, R. A., & Ohland, M.W. (2011). NSF's new data management policy: A conversation for engineering education research. Presented at NSF Engineering Education and Centers PI meeting, March 14, 2011.

Lord, S. M., Layton, R. A., & Ohland, M. W. (2011). Trajectories of electrical engineering and computer engineering students by race and gender. *IEEE Transactions on Education* **54**(4), November 2011, pp. 610–618.

Lord, S. M., Layton, R. A., Ohland, M. W., Brawner, C. E., & Long, R. A. (2014) A multi-institution study of student demographics and outcomes in chemical engineering. *Chemical Engineering Education* **48**(4), Fall 2014, 231–238.

Lord, S.M. (2010). Success in undergraduate engineering programs: A comparative analysis by race and gender. Invited talk at the *American Physical Society (APS) March Meeting*, Portland, OR, March, 2010.

Lord, S.M., Layton, R.A., & Ohland, M.W. (2015). Multi-institution study of student demographics and outcomes in electrical and computer engineering in the U.S.A. IEEE Transactions on Education, **58**(3), 141–150.

Love, B.J. (1993). Issues and problems in the retention of black students in predominantly white institutions of higher education, *Equity & Excellence in Education* 26(1).

Lundy-Wagner, V., Veenstra, C. P., Orr, M. K., Ramirez, N., Ohland, M. W., & Long, R. A. (2014). Gaining access or losing ground? Socioeconomically disadvantaged students in undergraduate engineering, 1994-2003. *Journal of Higher Education* 85(3), March/April 2014.

McNeil, Jacqueline, Matthew W. Ohland, and Russell A. Long, "Getting Better With Age: Older Students Achieve Higher Grades and Graduation Rates", Proc. 2014 IEEE/ASEE Frontiers in Education Conference, Madrid, Spain, October 22–25, 2014.

Min, Y., Zhang, G., Long, R. A., Anderson, T. J., & Ohland, M. W. (2011). Nonparametric survival analysis of undergraduate engineering student dropout. *Journal of Engineering Education*, **100**(2) 349–373, April 2011.

Mobley, F. C., Brawner, C. E., & Ohland, M. W. (2009). The South Carolina Merit Scholarship: Strategies used by engineering students to keep their LIFE Scholarship. *International Journal of Engineering Education*, **25**(6), pp. 1249–1256.

Muller, C. & Ellison, C. G. (2001). Religious involvement, social capital, and adolescents' academic progress: Evidence from the National Education Longitudinal Study of 1988. *Sociological Focus – Special Issue – Religion in America*, 34(2), 155–183.

Multiple-Institution Database for Investigating Engineering Longitudinal Development (MIDFIELD). (2015). Retrieved from https://engineering.purdue.edu/MIDFIELD

National Center for Educational Statistics (NCES). (2002). Retrieved from https://nces.ed.gov/statprog/2002/std4_2.asp

National Center for Educational Statistics Common Core of Data (NCES CCD). (2010). Retrieved from https://nces.ed.gov/ccd/

National Educational Longitudinal Study (NELS). (1988). Retrieved from http://nces.ed.gov/surveys/nels88/

National Educational Longitudinal Study (NELS). (2002). Retrieved from http://nces.ed.gov/surveys/els2002/

National Science Foundation Science and Engineering Indicators (NSF SEI). (2015). Chapter 2. Higher education in science and engineering. Retrieved from http://www.nsf.gov/statistics/seind14/index.cfm/chapter-2/c2h.htm

Ohland, M.W., & Anderson, T.J. (1999). Studying the Contribution of Programs at Eight Engineering Colleges toward Student Success. *Proc. Frontiers in Education (FIE) 1999*, San Juan, Puerto Rico, November 10-14, 1999.

Ohland, M.W., Lord, S.M., Layton, R.A. (2015). Student demographics and outcomes in civil engineering in the U.S. *Journal of Professional Issues in Engineering Education and Practice*. **141**(4), 1–7.

Ohland, M.W. (2005). Lessons from engineering education research: Ten things every faculty member should know. Keynote address at the Rose-Hulman Symposium, Rose-Hulman Institute of Technology, Terre Haute, Indiana, August 26, 2005.

Ohland, M.W. (2009). Using research to inform recruitment and retention efforts in engineering. EEC PI meeting, invited workshop February 1–2, 2009.

Ohland, M.W. (2012). Keynote address at the First-Year Engineering Experience Conference, Pittsburgh, PA, August 10, 2012.

Ohland, M.W., Brawner, C.E., Camacho, M.M., Long, R.A., Lord, S.M., & Wasburn, M.H. (2011). Race, gender, and measures of success in engineering education. *Journal of Engineering Education*, **100**(2), pp. 225–252, April 2011.

Ohland, M.W., Orr, M.K., Layton, R.A., Lord, S. M., & Long, R.A. (2012). Introducing 'Stickiness' as a versatile metric of engineering persistence. Proc. 2012 IEEE/ASEE Frontiers in Education Conference, Seattle, WA, October 3–6, 2012.

Ohland, M.W., Orr, M. K., Lundy-Wagner, V., Veenstra, C. P., & Long, R.A. (2012). 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., West Lafayette, IN: Purdue University Press, pp. 157–182.

Ohland, M.W., Sheppard, D., Lichtenstein, G., Eris, O., Chachra, D., & Layton, R.A. (2008). Persistence, engagement, and migration in engineering. Invited submission to a special issue of *Journal of Engineering Education,* **97**(3), 259–278.

Ohland, M.W., Lord, S.M., Long, R.A., Orr, M.K., & Brawner, C.E. (2016). Expanding Access to and Participation in the Multiple Institution Database for Investigating Engineering Longitudinal Development. NSF Award # 1545667, $4,010,978.00, 03/01/16 to 02/28/2021. http://www.nsf.gov/awardsearch/showAward?AWD_ID=1545667

Ohland, M.W., Zhang, G., Thorndyke, B. , and Anderson, T. J., Grade-point average, changes of major, and majors selected by students leaving engineering. Proc. Frontiers in Education 2004, Savannah, Georgia, November, 2004, paper 1478.

Orr, M. K., Lord, S. M., Layton, R. A., & Ohland, M. W. (2013). Student demographics and outcomes in mechanical engineering in the U.S. *International Journal of Mechanical Engineering Education* **42**(1), January 2014, 48-60, DOI: 10.7227/IJMEE.42.1.5.

Orr, M. K., Lord, S. M., Layton, R. A., Ramirez, N. M., Ohland, M. W. (2015). The future of aerospace engineering: Who will be the next generation of aerospace engineers? *Journal of Aerospace Information Systems,* **12**(4), 365–373.

Orr, M.K., Ramirez, N., & Ohland, M.W. (2011). Socioeconomic trends in engineering: Enrollment, persistence, and academic achievement. *Proc. American Society for Engineering Education 2011 Annual Conference*.

Orr, M.K., Ramirez, N., & Ohland, M. W. (2012). Using high school and district economic variables to predict engineering persistence. *Proceedings of the 2012 American Society of Engineering Education Annual Conference*, San Antonio, TX, June 2012.

Orr, M.K., Brawner, C.E., & Ohland, M.W. (2013). The effect of required introduction to engineering courses on major selection. *Proceedings of the 2013 American Society of Engineering Education Annual Conference*, Atlanta, GA.

Padilla, M.A., Zhang, G., Anderson, T. J., & Ohland, M. W. (2005). Drawing valid inferences from the nested structure of engineering education data: Application of a hierarchical linear model to the SUCCEED Longitudinal database. *Proc. Amer. Soc. Eng. Ed.*, Portland, Oregon, June 2005.

Ralston, K., Newman, C., Clauson, A., Guthrie, J., & Buzby, J. (2008). The National School Lunch Program: Background, trends, and issues. Economic Research Report Number 61. *US Department of Agriculture*.

Research findings contradict myth of high engineering dropout rate. Purdue News Release, August 4, 2009, Last accessed March 19, 2010, at http://news.uns.purdue.edu/x/2009b/090804OhlandEngineering.html

Ricco, G.D., Salzman, N., Long, R., & Ohland, M. W. (2012). Sectionality or why section determines grades: an exploration of engineering core course section grades using a hierarchical linear model and the Multiple-Institution Database for Investigating Engineering Longitudinal Development. *Proceedings of the 2012 American Society of Engineering Education Annual Conference*, San Antonio, TX, June 2012.

Rowan, B., Chiang, F., & Miller, R. J. (1997). Using research on employees' performance to study the effects of teachers on students' achievement. Sociology of Education, Vol. 70, No. 4 (Oct., 1997) , pp. 256–284. Retrieved from URL: http://www.jstor.org/stable/2673267.

Smith-Maddox, R. (1999). The social networks and resources of African American eighth graders: Evidence from the National Longitudinal Education Longitudinal Study of 1988. *Adolescence*, 34(133), 169–83.

U.S. Department of Education. National Center for Education Statistics (USDOE NCES). (2000). Analytic issues in the assessment of student achievement, NCES 2000–050 by David W. Grissmer and J. Michael Ross. Washington, DC. Retrieved from http://nces.ed.gov/pubs2000/2000050.pdf#page=197

Yoder, B. L. (2013). Engineering by the numbers. American Society Engineering Educuation. Washington, DC, USA. Retrieved from http://www.asee.org/papers-and-publications/publications/14_11-47.pdf

Zhang, G., Anderson, T. J., Ohland, M. W., Carter, R., and Thorndyke, B. (2004). Identifying factors influencing engineering student graduation and retention: A longitudinal and cross-institutional study. *Journal of Engineering Education* **93**(4), October 2004.

AUTHOR

Matthew Ohland is a Professor of Engineering Education at Purdue University. His research on the longitudinal study of engineering student development, team formation, peer evaluation, and extending the use of active and cooperative learning has been supported by the National Science Foundation and the Sloan Foundation. With his collaborators, he has been recognized with the best paper in the Journal of Engineering Education in 2008 and 2011 and in IEEE Transactions on Education in 2011 in addition to multiple conference best paper awards. Dr. Ohland is a Fellow of the American Society of Engineering Education and IEEE and has served on the IEEE Education Society Board of Governors (2007-2013), an Associate Editor of IEEE Transactions on Education, Chair of the Educational Research and Methods division of ASEE (2009–2011), and as a Program Evaluator for ABET. Dr. Ohland was the 2002-2006 President of Tau Beta Pi.

Russell Andrew Long is Director of Project Assessment and Managing Director of The Multiple-Institution Database for Investigating Engineering Longitudinal Development (MIDFIELD) at the Purdue University School of Engineering Education. He has extensive experience in large dataset construction and analysis, program review, assessment, performance based funding, and student services in higher education. One of his greatest strengths lies in analyzing data related to student learning outcomes and, therefore, to improving institutional effectiveness. His work with MIDFIELD includes research on obstacles faced by students that interfere with degree completion and how institutional policies affect degree programs.