The Creation of the Multiple-Institution Database for Investigating Engineering Longitudinal Development (MIDFIELD)

Matthew W. Ohland, Guili Zhang, Brian Thorndyke, and Timothy J. Anderson
General Engineering, Clemson University / Educational Statistics, Physics, and Chemical Engineering, University of Florida

Abstract

A longitudinal database containing student records from each of the universities participating in the Southeastern University and College Coalition for Engineering Education (SUCCEED) has been under continuous development since 1996. The database compiles records from 1987 through the present. An extensive body of research has been conducted using this database.

Technological advancements have made it possible to expand the scope of the data included in the database, and thus expand the studies that can be conducted. The partner universities have agreed to contribute transcript information—the courses that students have taken and the grades they have earned in them. With these data from nine universities placed in a common format, exciting research questions can be addressed.

This paper discusses the development of the original database, plans for the design of the MIDFIELD database, and expected uses of this valuable resource.

The SUCCEED Longitudinal Database

A longitudinal database is under continuing development by the SUCCEED Coalition with records for undergraduate students from 1987 to present (2002 data are being added) at each of the SUCCEED institutions:

- Clemson University
- Florida A&M University
- Florida State University
- Georgia Institute of Technology
- North Carolina A&T State University
- North Carolina State University
- University of Florida
- University of North Carolina at Charlotte
- Virginia Polytechnic Institute and State University

Data from all nine universities have been placed in a common format, so the SUCCEED Longitudinal Database can be used for cross-institutional studies. Descriptions of the database design and assembly are available, and the fields presently included in the database are included later. The database is a unique resource that has been studied extensively. Highlights of publications related to the present work include

- studies of retention and performance of students in a number of special programs implemented in the SUCCEED Coalition, including various freshman programs, integrated curricula, an entrepreneurship program, minority engineering programs,
• descriptive studies comparing the characteristics of engineering students with each other and with non-engineering students
• analysis of the paths taken by students leaving engineering
• coordinating a quantitative study of retention of women in engineering with a qualitative study of the climate performed at SUCCEED institutions

These publications, and others in review and in development, are indicators of the value of this database resource. One of the more interesting studies indicates that students who leave engineering tend to change into a major according to their grade-point average at the time of attrition. This certainly has implications for retention efforts and advising.

The Opportunity to Expand the SUCCEED Longitudinal Database

Negotiations initiated in 2002 with the SUCCEED partner institutions resulted in a new partnership to form the Multiple-Institution Database for Investigating Engineering Longitudinal Development (MIDFIELD). Prior to this new partnership, the participating institutions had already committed to providing supplementary data annually to the SUCCEED Longitudinal Database until three years after the end of the SUCCEED Coalition (NSF funding terminated in 2003). The database in its current configuration has significant value to the engineering education community. The value will be improved significantly through the addition of course enrollment and grade data for the entire population. The addition of course and grade data will make possible a broader range of research questions.

The MIDFIELD Model of Student Pathways in Engineering

To study course sequencing and student pathways, it is necessary to model the engineering curriculum in such a way that different engineering curricula can be interpreted through a common framework. On the surface, this might seem to cast aside the variation needed to study these issues. This common framework is needed to even begin analysis of the curriculum of the various schools. In this case, the variation will be provided by the students themselves, who will take myriad pathways through the curriculum as a result of differences in their high school preparation, scheduling preferences, educational objectives, and other factors.

This common framework is the MIDFIELD Model of Student Pathways in Engineering, which generalizes the paths a student might take to graduate in engineering and the paths a student might take before leaving engineering. The model, shown in Figure 1, creates a common framework by defining checkpoints in the engineering curriculum as stages in the model. The model can be expanded to study a particular part of the curriculum in greater detail—such as specifically where the students who leave engineering go. In studying data from multiple institutions, this broad model is most applicable, because the greater the detail of the model, the more divergent the institutions, their curricula, and their policies become.
Figure 1: The MIDFIELD Model of Student Pathways in Engineering

The model links the inputs of the engineering education system, first-time-in-college and transfer students matriculating to engineering programs to the outputs of that system, students who graduate in engineering programs. The system has byproducts (students who do not persist in engineering) and feedback loops (students who retake a course in order to continue in engineering). As the model shows, there are various points at which students may choose not to persist in engineering in spite of the fact that they are on a successful path. In this model, student performance in the first college math course they take is singled out both because of its recognized role in the development of a student’s confidence in engineering and because of the significant role that choice of course plays in determining a student’s pathway through the engineering curriculum.\textsuperscript{18-24} The “engineering core” courses are generally well defined and include a three-course sequence in calculus, a course in differential equations, a two-course sequence in physics, and a two-course sequence in chemistry. As a result of the work of the Coalition, the SUCCEED institutions all have some engineering content during the freshman year, so those courses would also be included in the engineering core. “Engineering gateway” courses are the first courses that a student takes that define the discipline that they have chosen to major in, and in which the student must succeed in order to take further coursework within the discipline. Frequently, such gateway courses have high failure rates. Some disciplines will have more than one discipline-specific course that a student takes as they begin coursework in their discipline—in this case, the combination of those courses will be treated as a unit.

\textbf{Descriptive statistics determine the fraction of students in each path.} A significant contribution of MIDFIELD will simply be to identify the number of students that take each path after matriculating in engineering. Qualitative research studies published in the past decade have discussed why students leave engineering and the sciences.\textsuperscript{25,26} Descriptive statistics determining the relative size of the model pathways will supplement those qualitative studies by identifying how many students are leaving and when. Institutional differences in gender effects will be interpreted in light of the results of climate studies previously supported by SUCCEED.\textsuperscript{27}
Focus group populations. It has long been understood that we cannot learn why students are leaving engineering if we study only those students who are retained in engineering. To look only at retained students is myopic and does nothing to enhance the diversity of engineering—just the opposite, it tends to promote the continuing development of engineers that are drawn from an even narrower group. In the qualitative research supporting this descriptive study, we will pay special attention to those that are at the interface of retention. There are two groups of interest at this interface: (1) students who persist in engineering in spite of the need to repeat courses that they have failed (sometimes multiple times) in order to remain in engineering and (2) students who do well in the engineering core and gateway courses (who would seem to be “over the hump”), but leave engineering anyway. Focus groups will be drawn from the four populations shown in Table 1:

<table>
<thead>
<tr>
<th>Pass at model checkpoints</th>
<th>Fail at model checkpoints</th>
</tr>
</thead>
<tbody>
<tr>
<td>Persist in engineering</td>
<td>Students who do well and persist in engineering as planned</td>
</tr>
<tr>
<td>Leave engineering</td>
<td>Students who do well in engineering courses, but choose to leave engineering</td>
</tr>
</tbody>
</table>

Retaining institutional variability. Institutional and disciplinary differences that have been artificially masked by the common framework of the model may result in differences in the proportion of students following each path. For example, if a particular engineering discipline has a significant problem with retention in their gateway course(s), the path representing successful completion of that gateway course will be disproportionate when compared to the path for the same discipline at other institutions. In this way, the variability of the institutions is still present, and makes it possible to study the effect of those gateway courses.

Questions addressed by frequency and descriptive statistics. The frequency and descriptive statistics of student pathways provide a broader picture of issues that are more commonly performed on students of a single institution. These will help identify general trends across institutions on the issues of concern listed below, which are likely to suggest further qualitative and quantitative studies:

- How likely are students to succeed in their first math course on campus?
- How likely is a student to stay in engineering if they succeed in their first math course?
- Does performance in the first math course predict performance in other math courses?
- How likely are students to succeed in the engineering core courses?
- How many students complete the engineering core and yet decide to leave engineering?
- How prevalent are gender and ethnicity differences and are they changing over time?
- How likely are students to successfully complete engineering gateway courses?
- How many students complete the engineering gateway courses and yet leave engineering?
- When students leave engineering after doing well, what majors do they choose?
- How well do they perform in these alternate majors?
- When students leave engineering after performing poorly, what majors do they choose?
• How well do they perform in these alternate majors?
• Is there a difference in the alternate majors that are chosen by the students who do well compared with the students who do poorly in the engineering core and gateway courses?
• Is there a difference in the characteristics of the students who choose alternate majors controlling for their performance in the engineering core and gateway courses?

Example of a Special Study: The Influence of State Scholarships

The MIDFIELD database will have application well beyond the mapping of student pathways as described above. Special studies focusing on particular issues will enhance the value of the database. Since the adoption of a state scholarship program in South Carolina, Clemson’s General Engineering advisors (advising mostly engineering freshmen) observe that students frequently make academic decisions to protect their scholarship. Exit surveys and advising records are expected to indicate that this is pervasive. Using the MIDFIELD database, we will study scholarship status as a predictor of engineering retention. Since scholarship status is directly related to GPA, the study design seems confounded, but institutional diversity provides sufficient variation—Florida, Georgia, and South Carolina (5 institutions) have state-based scholarship support (Bright Futures, Hope, and Life scholarships, respectively), while North Carolina and Virginia (4 institutions) do not. If scholarship protection strategies have affected engineering retention, this may be revealed by a transition in retention rates at the FL, GA, SC states at the scholarship threshold, but no such transition should be present in data from the NC and VA schools. To establish the causality of the quantitative findings, campus liaisons will provide names and contact information for students with GPA’s near the scholarship criterion who stayed in and left engineering so that those students may be asked to participate in interviews to probe more deeply into their reasons for leaving.

The Generalizability of Results of Studies at MIDFIELD Institutions

According to the Engineering Workforce Commission, SUCCEED schools enroll over 28,000 engineering undergraduates, award 1/12th of U.S. engineering degrees, a higher percentage of all U.S. engineering degrees awarded to women (1/10), and a significant portion (1/4) of U.S. engineering degrees awarded to blacks; this sample size helps ensure that the results can be applied to other institutions.\textsuperscript{28,29} Summary statistics for participating institutions are compared to national averages in Tables 3 and 4.

The comparison favors generalizability with the notable exception of the Historically Black College and University (HBCU) participants (Florida A&M University and North Carolina A&T State University). The total enrollment is three times the national average, but the mean is likely to be a poor measure of central tendency of engineering enrollment statistics, since there are many schools with low enrollments. These low values are presumed to skew the distribution (the actual distribution is unavailable to confirm this). Racial composition generally matches national averages, with the expected exception of the HBCU participants. The only outlier is FSU, which likely has a larger percentage of black students because it shares a College of Engineering with Florida A&M. With the exception of the HBCU participants, gender composition is also typical. We suspect that the HBCUs attract more female students because relevance or the curriculum to
society, a strong community, and other features that particularly improve the learning environment for minority students also correlate positively with the success of women. A qualitative study will be designed to investigate this. In the worst case, the results should be generalizable for larger schools, which make the findings applicable to a significant majority of US engineering students. SAT Quartiles are not available for a national sample, and SAT statistics for engineering majors only are not available, so comparison is difficult.

### Table 3. Undergraduate Engineering Enrollment, Gender Composition, and SAT Quartiles of Partner Institutions

<table>
<thead>
<tr>
<th>Institution</th>
<th>Engineering enrollment</th>
<th>Gender composition</th>
<th>SAT Quartiles+</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td>Female</td>
<td>Male</td>
</tr>
<tr>
<td>Clemson</td>
<td>2789</td>
<td>514</td>
<td>18%</td>
</tr>
<tr>
<td>FAMU</td>
<td>729</td>
<td>494</td>
<td>68%</td>
</tr>
<tr>
<td>FSU</td>
<td>1266</td>
<td>272</td>
<td>21%</td>
</tr>
<tr>
<td>Ga Tech</td>
<td>6336</td>
<td>1427</td>
<td>23%</td>
</tr>
<tr>
<td>NCAT</td>
<td>1408</td>
<td>513</td>
<td>30%</td>
</tr>
<tr>
<td>NC State</td>
<td>5551</td>
<td>994</td>
<td>18%</td>
</tr>
<tr>
<td>U. Florida</td>
<td>4592</td>
<td>956</td>
<td>21%</td>
</tr>
<tr>
<td>UNCC</td>
<td>915</td>
<td>109</td>
<td>12%</td>
</tr>
<tr>
<td>Va Tech</td>
<td>4667</td>
<td>758</td>
<td>16%</td>
</tr>
<tr>
<td>Natl. Av.</td>
<td>1079</td>
<td>213</td>
<td>20%</td>
</tr>
</tbody>
</table>

*Average in each category based on 340 engineering schools
+ Includes all majors
++ Florida state gender composition estimated from statistics that included graduate students

### Table 4. Undergraduate Engineering Enrollment and Racial Composition of Partner Institutions

<table>
<thead>
<tr>
<th>Institution</th>
<th>ENGR enrmt.</th>
<th>Racial Composition</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td>Black</td>
</tr>
<tr>
<td>Clemson</td>
<td>2789</td>
<td>2274</td>
</tr>
<tr>
<td>FAMU</td>
<td>729</td>
<td>15</td>
</tr>
<tr>
<td>FSU</td>
<td>1266</td>
<td>767</td>
</tr>
<tr>
<td>Ga Tech</td>
<td>6336</td>
<td>4273</td>
</tr>
<tr>
<td>NCAT</td>
<td>1408</td>
<td>98</td>
</tr>
<tr>
<td>NC State</td>
<td>5551</td>
<td>5174</td>
</tr>
<tr>
<td>U. Florida</td>
<td>4592</td>
<td>3179</td>
</tr>
<tr>
<td>UNCC</td>
<td>915</td>
<td>703</td>
</tr>
<tr>
<td>Va Tech</td>
<td>4667</td>
<td>3490</td>
</tr>
<tr>
<td>Natl. Av.</td>
<td>1079</td>
<td>730</td>
</tr>
</tbody>
</table>

*Average in each category based on 340 engineering schools
+ NC State data for “White” includes White/Other
Project management and data integrity

**Project management.** The management of the database itself will be enhanced by separating the distinct functions of database maintenance and database analysis. The former will be managed by a team at Clemson University and the latter by a team at the University of Florida. This approach will help ensure that the dataset is validated before analysis. The project teams’ familiarity with the partner institutions and their data management practices in the past 15 years will be a significant asset in understanding how to correctly interpret institutional differences. Changes in file formats, changes to ensure Y2K compatibility, changes in academic policies, and even changes from quarter-hour system to a semester-hour system have already been accounted for in the analysis of the existing database. Validation of the database, which has previously relied on review of time series of enrollment, graduation, and other statistics for inconsistencies, will now generate standard reports for each institution providing data. These reports will be reviewed and approved by institutional representatives that make up an advisory board.

**Data integrity.** A database of this scope is not without integrity issues, but with the help of talented personnel at partner institutions, the dataset is surprisingly complete. Three institutions currently lack SAT scores, but it is expected that these can be secured along with the startup course and grade dataset of the MIDFIELD project. In previous studies, where regressions included SAT scores, those institutions used a different model that did not include the SAT scores as a predictor. Having some schools on a quarter system and others on a semester system as partners has been overcome using standard formulas to convert quarter hours into semester hours. Some institutions could not provide an accurate flag indicating a student’s participation in a cooperative education program, so this has been managed by counting semesters-to-graduation rather than time-to-graduation. This is becoming a greater concern as the research team has recently been studying the characteristics of students in cooperative education programs. Fortunately, the institutions that do not have an accurate co-op flag have a course indicating co-op participation, which will be included with the new data added.

**Letters of Support and Institutional Review Board Approval**

The SUCCEED Longitudinal Database was created with the cooperation of high-level administrators at each institution—Deans of Engineering, Institutional Research Directors, Registrars, Chancellors, and others. Letters of support are on file that indicate that data will be provided to the project for three years after the completion of SUCCEED Cooperative Agreement in 2002. Since MIDFIELD is a significant expansion of the project, new letters of support were requested and received. Clemson’s Institutional Review Board has already approved the protocol for the collection and study of data in the original database, and IRB personnel are assisting in the development of a revised protocol for MIDFIELD.

**Conclusions**

The MIDFIELD database is a powerful tool in learning more about the behaviors of students who matriculate to engineering programs—including the paths they take, the choices they make, what majors they change to and in what major and when they graduate. It is important to note
that the results of quantitative studies require the application of considerable judgment including the search for alternative explanations for observed relationships. The best quantitative results are those which in some way corroborate the results of qualitative studies. Indeed, qualitative study is frequently the only way in which the results of quantitative analysis can be understood. This database, as is true of any database, cannot provide a complete picture of student development.

Author Biographies

MATTHEW OHLAND is Assistant Professor in General Engineering at Clemson University and is the President of Tau Beta Pi, the national engineering honor society. He received a Ph.D. in Civil Engineering with a minor in Education from the University of Florida in 1996. He previously served as Assistant Director of the NSF-sponsored SUCCEED Engineering Education Coalition. His research is in freshman programs and educational assessment.

BRIAN THORNDYKE is a Ph.D. candidate in the Quantum Theory Institute, Department of Physics, University of Florida. He received a M.Sc. in High Energy Physics at the University of Montreal, and an M.S. in Computer Science at the University of Florida. His research interests involve computational methods applied to the dynamics of mixed quantum/classical molecular systems.

GIULI ZHANG is a Ph.D. candidate in Educational Research and Statistics, Department of Educational Psychology, University of Florida. She received a B.A. in British and American Language and Literature at Shandong University, China, and a M.Ed. in English Education at Georgia Southern University. She has published extensively and has won numerous awards at the national and regional level in the area of educational research in China. She teaches Measurement and Assessment in Education at the University of Florida. Her research interests involve applied quantitative research designs, categorical data analysis, and structural equation modeling.

TIM ANDERSON is Associate Dean for Research and Graduate Programs in the College of Engineering, and Professor in the Department of Chemical Engineering, University of Florida. He received a Ph.D. at the University of California-Berkeley in 1979. His research interests include electronic materials processing, thermochemistry and phase diagrams, chemical vapor deposition, bulk crystal growth and advanced composite materials.

References


Proceedings of the 2004 American Society for Engineering Education Annual Conference & Exposition
Copyright © 2004, American Society for Engineering Education
   http://www2.acs.ncsu.edu/UPA/fastfacts/trendenr.htm
   http://apir01nt.uncc.edu/irmainpage2/fb02/fb02036-s.html
38. Virginia Tech Factbook statistics. (2003). Used by permission of Bevlee Watford, Associate Dean, College of 
   Engineering, May 27, 2003
   Foundation.
   <http://www.usnews.com/>