How Science Course Performance Influences Student Retention - A Statistical Investigation

Guili Zhang, Brian Thorndyke, Matthew W. Ohland, and Timothy J. Anderson

Educational Psychology Department, University of Florida / Department of Physics, University of Florida / Department of General Engineering, Clemson University / Department of Chemical Engineering, University of Florida

Abstract

Identifying science courses that influence retention of engineering students should be useful in suggesting approaches to improving student success by focusing the attention of educational reforms on the areas of greatest need. The purpose of this study is to statistically identify those required science courses in engineering degree programs that have a significant influence on retention and estimate the magnitude of their impact. We draw our inferences from a database of all engineering students at the University of Florida between Spring 1996 and Spring 2002. Defining retention as either graduation or current enrollment in the final year of the database, a series of logistic regression models are formulated to estimate the effect on retention in engineering by core science courses in three major areas: mathematics, physics and chemistry. Odds ratios on retention are reported and rank ordered as an indication of the strength and relative importance of each course's impact, and we find that the student performance in only a fraction of the core courses has predictive value on engineering retention. These results raise questions about the relationship of the core courses to the later curriculum, and suggest that curriculum development and academic advising should reflect the variation within the science core and resist the temptation to treat the core uniformly.

I. Introduction

Most engineering programs require students to take a set of first- and second-year science courses as core requisites, and often these core courses are independent of the engineering specialty. At the University of Florida, we require all engineering undergraduate students to successfully pass a set of science courses chosen from mathematics, chemistry and physics. Implicit in this curriculum requirement is that these courses are fundamentally important to students' success in engineering, as they develop the basic tools of mathematics and scientific

Table 1: List of core science courses taken by engineering students at UF

Group	Abbreviation	Course Description
Math	Calc1	Analytical Geometry and Calculus 1
	Calc2	Analytical Geometry and Calculus 2
	Calc3	Analytical Geometry and Calculus 3
	Calc1Hon	Honors Analytical Geometry and Calculus 1
	Calc2Hon	Honors Analytical Geometry and Caluclus 2
	Calc3Hon	Honors Analytical Geometry and Calculus 3
	DiffEq	Differential Equations
	CompSci	Computer Programming for Engineers
Physics	Phys1	Physics with Calculus 1
	Phys2	Physics with Calculus 2
	Phys1Hon	Honors Physics with Calculus 1
	Phys2Hon	Honors Physics with Calculus 2
	Phys1Lab	Physics with Calculus 1 Laboratory
	Phys2Lab	Physics with Calculus 2 Laboratory
Chemistry	ChemIntro	Introduction to Chemistry
	Chem1	General Chemistry 1
	Chem2	General Chemistry 2
	Chem1Hon	Honors General Chemistry 1
	Chem1Lab	General Chemistry 1 Laboratory
	Chem2Lab	General Chemistry 2 Laboratory

analysis that will be needed throughout the engineering degree. It stands to reason that a strong performance in the science core should contribute significantly to a student's understanding and performance in advanced engineering courses, leading to a greater probability that the student will remain in engineering and graduate.

In previous work,[1-4] we performed cross-institutional studies to determine the relationship of various factors to the likelihood of retention of engineering undergraduates, where retention refers to the graduation or enrollment in engineering as of the last entry in the database. Our results demonstrated that SAT math scores, SAT verbal scores, high school GPA, gender, ethnicity and citizenship are all significantly related to the retention rates of engineering students. However, none of these studies included course-level data, so questions related to specific course-related factors could not be answered.

In this work, we extend our database to include course-level data in core science courses while at the same time limiting the study to the University of Florida, allowing us to track the individual grades of every student as they progress toward their engineering degree at a single institution.

By fitting results to logistic regression curves, we are able to determine quantitatively the relationship between student performance in the core courses and student retention, as well as statistically differentiate the effect of these courses. These analyses provide valuable insight into the impact of the science core within the engineering program, and provoke questions relating to both curriculum development and student advising.

II. Data Collection

The database used in this study derives from the longitudinal database (LDB) previously maintained by the SUCCEED collaboration,[5-11] which gathered complete undergraduate demographic, term and graduation data for all undergraduate students between 1987 and 2002. Through the LDB, we were able to discover many factors influencing retention, but since course data was unavailable, these results were limited to demographic factors and overall term performance. Building on the LDB, we augmented the database with grade information on twenty core science courses at UF from Spring 1996 through 2002. Grade point values (GPVs) were obtained for all UF students in these courses, which allowed us to model the probability of student retention as a function of GPV in a given course. We list these core science courses in Table 1.

The LDB contains both first-time-in-college (FTIC) and transfer students. In the present work, we limit our study to FTIC students only, and also limit the grade range considered.
While failure in a core course obviously negatively affects retention, since the student may not be permitted to continue registration until the grade is corrected, we are more interested in the predictive nature of the GPV for successful students. Put another way, given a student has passed a given course, we would like to estimate that student's chances of retention depending on their grade in that course. To this end, we included only GPVs which are at least 2.0, the passing GPV for core science courses UF. In spite of these restrictions, our data from the core science courses encompassed over 27,500 grades from engineering students and 8500 grades from non-engineering students.

III. Analysis

A. Statistical Methods

We are interested in the statistical relationship between student performance in each course and student retention. To this end, we have modeled retention vs. performance along logistic regression curves, which are suitable for analyzing data where the independent variable takes on a continuous range of values but the dependent variable is discrete. From these logistic analyses we are able to infer the probability of retention as a function of course grade, as well as the statistical significance of the results. We present an overview of the modeling procedure in this subsection, and refer to Neter et al.[12] for further mathematical details.

In order to build a model of student retention (labeled 'Y') as a function of GPV (labeled 'X'), we need to take into account the dichotomy of our dependent variable: Either the student graduates/stays in engineering (Y = 1) or they fail/leave engineering (Y = 0). On the other hand, the independent variable is continuous: The student's GPV takes on the domain $X \in [2.0, 4.0]$. This kind of relationship can be modeled by a sigmoidal function,

$$\pi(X) = \frac{\exp(\beta_0 + \beta_1 X)}{1 + \exp(\beta_0 + \beta_1 X)},$$

where $\pi(X)$ is the probability of retention for GPV = X. The sigmoidal function is appropriate, because it shows asymptotes $\pi(-\infty) = 0$ and $\pi(\infty) = 1$, and expresses linearity in the region around $\pi = 0.5$.

Our goal is to determine the constants β_0 and β_1 such that $\pi(X)$ is the best fit to the data,

$$E(Y_i) = \frac{\exp(\beta_0 + \beta_1 X_i)}{1 + \exp(\beta_0 + \beta_1 X_i)},$$

where $E(Y_i)$ is the expectation of retention for GPV = X_i based on the student data. Since β_0 and β_1 cannot be expressed in closed form, a maximum likelihood analysis is numerically performed to produce the best fit. Thre resulting function $\pi(X)$ can then be used to estimate the effect of GPV on retention.

Specifically, we begin with the so-called odds (O) of retention,

$$O(X) = \frac{\pi(X)}{1 - \pi(X)},$$

an indicator similar to probability of retention, but without an upper bound. We are interested in evaluating the effect of improved grades in a course on the odds of retention, which leads us to consider the odds ratio (OR),

$$OR = \frac{O(X+1)}{O(X)} = \frac{\pi(X+1)[1 - \pi(X)]}{\pi(X)[1 - \pi(X-1)]}.$$

The OR gives us the multiplicative increase in odds of retention resulting from a 1-point increase in GPV. For example, if a course has an OR = 1.26, then a student increases his odds of retention by 26% for every 1-point increase in GPV.

Substituting the expression for π and simplifying, we find that the OR can be expressed simply as a function of β_1,

$$OR = \exp(\beta_1) .$$

In what follows, we show results of logistic regression fits for each core science course, computed using the SAS 9.0 package. Indicated are the ORs, as well as their 95% Wald confidence intervals (CIs).

B. Results: Influence on Retention

We fit the student data to logistic regression curves for each of the core science courses listed in Table 1. For those courses whose GPVs were found to have significant effects on student retention in engineering, we present their odds ratios and rank them in order of importance in Table 2. It is remarkable that out of the 20 courses considered, only 6 of them were significant. For the remaining 14 courses, a student's GPV in the course did not statistically reflect their probability of retention.

In terms of magnitude of effect, the General Chemistry 1 Laboratory had the most predictive value, with an OR = 1.59. This means that, for example, an engineering student with a GPV = 3.0 (a B) in this course has a 59% greater odds of retention than a student with GPV = 2.0 (a C). On the other end of the spectrum, the Physics with Calculus 1 Laboratory was the least predictive of the significant courses, with an OR = 1.21. We also note that there is a clear grouping in the rankings, with the chemistry courses showing the most influence, followed by the math courses and finally the physics course.

On the other hand, the 14 other core science courses showed no significance to retention. This means that, for example, a student in Differential Equations with a maximum GPV = 4.0 would have no statistically increased odds of retention over another student with a barely passing GPV = 2.0. It is very interesting, and indeed counterintuitive, that the vast majority of the core courses fell into this category. A possible explanation for the insignificance of the honors courses is that students entering these advanced classes are somewhat above average, and will all likely succeed, so that differences in their GPVs are not important to retention. It is more difficult to explain the lack of significance in Differential Equations, the traditional Physics with Calculus 1 & 2, Introductory Chemistry and the standard General Chemistry 2, and finally Computer Programming for Engineers.

At the University of Florida, these courses are required by all engineering students during their first two years in the undergraduate program, with the notion that they are fundamentally important to their overall success in engineering. From the results of this study, it appears that their connection to the later curriculum is not so obvious with regard to student retention. This does not imply that these courses are not important to engineering students, but the findings do open many questions about the precise way in which they fit into a successful engineering program. For instance, it may be possible to introduce more flexibility into the curriculum by changing the status of certain science courses from prerequisite to co- or even postrequisite with junior- or senior-level engineering courses, without negatively affecting student success.

Table 2: Courses ranked according to the significance of student GPV to retention in engineering. Courses which did not have any predictive effect are not listed.

Rank	Course	Odds Ratio	95% Wald CI
1	Chem1Lab	1.590	[1.308, 1.934]
2	Chem2Lab	1.449	[1.182, 1.776]
3	Chem1	1.447	[1.287, 1.626]
4	Calc2	1.401	[1.228, 1.598]
5	Calc1	1.318	[1.138, 1.527]
6	Phys1Lab	1.214	[1.045, 1.412]

There are also implications for academic advising during the freshman and sophomore years. Early in a student's program, recognition of the core courses with primary influence on retention should suggest the areas which require the most remedial focus for incoming freshmen with prerequisite deficiencies. In a similar spirit, once a student has completed one or two years of coursework, it may be important to compute their grade point average from subsets of core courses statistically significant to retention (rather than dividing between the full core set and electives, as is commonly done), in order to accurately assess their probability of success and recommend appropriate supplementary study and course sequences to complete their degree.

C. Student Performance

Given a clear division between the core science courses, we thought it would be interesting to look at the overall GPVs in each course, and in particular to contrast the performance of the engineering students with non-engineering students. The results are presented in Tables 3 and 4.

In Table 3, we see the student performance in each of the 6 predictive courses. While they have been listed in rank of predictive power, there is no correlating order among the engineering student GPVs in the courses. The average engineering GPVs range from 3.03 to 3.60, with no particular pattern. On the other hand, the difference in performance between engineering and non-engineering students is substantial. In every case, the engineering students perform better on average than their non-engineering counterparts, with at least a 99% confidence level.

In Table 4, we present the student performance in each of the remaining 14 non-predictive courses. While there is not any way of ranking these courses in terms of predictive power, we note that the average engineering GPVs have even a larger range than found for the predictive courses, spanning 2.91 to 3.82. Similar to the predictive courses, however, is a disparity in

Table 3: Student performance in predictive courses.

Course	N	Mean	Stdev	N	Mean	Stdev
	(Engineering Students)			(Non-Engineering Students)		
Chem1Lab	3093	3.66	0.51	1304	3.45	0.59
Chem2Lab	1018	3.60	0.62	549	3.51	0.60
Chem1	2902	3.03	0.72	411	2.75	0.69
Calc2	1987	3.21	0.73	822	2.88	0.76
Calc1	1969	3.19	0.75	991	2.88	0.78
Phys1Lab	2113	3.22	0.58	942	2.99	0.54
Average	13082	3.30	0.69	5019	3.11	0.72

Table 4: Student performance in non-predictive courses.

Course	N	Mean	Stdev	N	Mean	Stdev
	(Engineering Students)			(Non-Engineering Students)		
Calc3	2030	3.30	0.70	482	3.18	0.76
DiffEq	2052	3.34	0.70	404	3.09	0.79
Calc1Hon	164	3.45	0.65	51	3.26	0.59
Calc2Hon	631	3.41	0.65	137	3.11	0.84
Calc3Hon	231	3.37	0.71	45	3.19	0.72
CompSci	791	3.16	0.70	67	2.98	0.78
Phys1	1968	2.91	0.72	416	2.67	0.68
Phys2	1872	3.01	0.69	355	2.64	0.68
Phys2Lab	2246	3.21	0.59	548	2.95	0.61
Phys1Hon	50	3.43	0.70	20	3.85	0.49
Phys2Hon	30	3.53	0.51	12	3.46	0.50
ChemIntro	1338	3.17	0.75	675	2.80	0.71
Chem1Hon	39	3.82	0.39	16	2.72	0.88
Chem2	1164	3.12	0.72	315	2.81	0.68
Average	14606	3.18	0.71	3543	2.91	0.74

grades between the engineers and non-engineers: In every course except Honors Physics 1, the engineers outperform the non-engineers. However, the differences between engineers and non-engineers in the Honors classes are not as statistically significant (reaching only 95% confidence in some cases), while the differences in courses like Calculus 3 and Differential Equations reach the 99% confidence level.

From the GPV results of Tables 3 and 4, it is clear that engineers generally perform better than non-engineers in the core science courses. However, it seems that neither the average GPV in a given course, nor the disparity in GPV between the engineers and non-engineers, can serve as an indicator of the predictive value of that course to engineering retention.

IV. Conclusions

Building on a database of demographic, term and graduation data for all undergraduate students at the University of Florida, we added course-specific data for twenty core science courses required by all engineering undergraduate students. Through a series of logistic regression analyses, we were able to model the predictive relationship between student performance in these courses and retention in engineering.

We found that only six of the core science courses showed a predictive effect, in that a student's grade in each of these courses was positively correlated with his odds of retention in the engineering program. Among these predictive courses, the most influential were General Chemistry labs and the first General Chemistry lecture. Calculus 1 and 2 fell behind the chemistry courses, while the first Physics lab was the only course among the Physics group which showed predictive value.

Among the remaining fourteen core courses, a student's grade had no statistically significant effect (at the 95% confidence level) on retention in engineering. This is surprising, since one would assume that all of the core courses would be very important to student success, and that high performance in these courses would necessarily predict retention. This was not the case for these remaining courses, although some of them, such as the honors classes, might be explained by noting that students entering honors classes will generally be better students, who will most likely succeed regardless of their grade in the course. This explanation does not hold, however, for basic courses like the Physics lectures, Introductory Chemistry and the second Chemistry lab, Differential Equations, and Computer Programming for Engineers.

We also looked at the relative performance in each course between engineers and their non-engineering counterparts. We found that in virtually every case, for both predictive and non-predictive courses, the engineers had better grades, usually with over a 99% confidence level. However, there appeared to be no correlation between either the average grades, or disparity in grades, and the ability of a course to predict engineering retention.

The results of this study open some interesting questions about the connection of core science courses to the rest of the engineering curriculum. We have shown that within the University of Florida, a student's grade in a core science course is often not a predictor of retention in engineering. These core courses are normally prerequisites for more advanced and specialized engineering courses, but in light of these findings, it may be possible to introduce more flexibility into the curriculum and allow students to take some of the core courses (presumably, the ones which show no correlation with student retention) later in their program, without any negative affect on student success. There are also implications for student advising, where it might be more effective to focus on a select subset of core science courses in recommending remedial work and course sequencing, rather than treating the science courses as a single, undivided set.

For the near future, we are looking into performance/retention relationships within each engineering sub-discipline, as it is very likely that different core science courses are important to different subfields. It would also be interesting to augment the study with qualitative investigations into the differences between the predictive and non-predictive courses, in an attempt to determine the relative importance of factors like course curricula and instructor effectiveness. In the longer term, we are also involved in a multi-institutional collaboration (MIDFIELD) to develop a longitudinal database including course-level data for all undergraduate courses across several institutions,[13] which would allow us to investigate the role of core curricula in student success for a wide range of engineering programs, and perhaps generalize our results to institution-independent conclusions.

References

[1] Zhang, G., B. Thorndyke, M. Ohland and T. Anderson (2003) "Comparing the Performance of Chemical Engineering Students with Other Disciplines," Chemical Engineeriing Education (submitted, October 2003).

[2] Zhang, G., T. Anderson, M. Ohland, R. Carter and B. Thorndyke (2003) "Identifing Factors Influencing Engineering Student Graduation and Retention: A Longitudinal and Cross-Institutional Study," Journal of Engineering Education (submitted, July 2003).

[3] Anderson, T., R. Carter, M. Ohland, B. Thorndyke, and G. Zhang (2002) "Identifying Factors Influencing Engineering Student Retention through a Longitudinal and Cross-Institutional Study Using Quantitative and Qualitative Methods," *Proc. Amer. Soc. Eng. Ed.*, Montreal, Canada, June 2002.

[4] Zhang, G., T. Anderson, M. Ohland, R. Carter, and B. Thorndyke (2002) "Identifying Factors Influencing Engineering Student Graduation and Retention: A Longitudinal and Cross-Institutional Study," *Proc. Amer. Soc. Eng. Ed Southeast.*, Gainesville, FL, April 2002, Session 2793.

[5] Ohland, M.W., and T.J. Anderson (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.

[6] Ohland, M.W., and T.J. Anderson (1999) "Studying the Contribution of Programs at Eight Engineering Colleges toward Student Success," *International Conference on Engineering Education 1999*, Ostrava and Prague, Czech Republic, August 8-14, 1999, published in proceedings on CD-ROM, paper 436, 9 pages (indexed by the ERIC database with accession number ED452047).

[7] Carson, Lewis, "SUCCEED Quantitative Evaluation," *The Innovator*, No. 8, Fall 1997, SUCCEED Engineering Education Coalition, University of Florida, Box 116134, Gainesville, FL 32611-6134.

[8] Ohland, M.W., and S.A. Rajala, and T.J. Anderson (2001) "SUCCEED-Sponsored Freshman Year Engineering Curriculum Improvements at NC State: A Longitudinal Study of Retention," *Proc. Amer. Soc. Eng. Ed.*, Albuquerque, NM, June 2001.

[9] Ohland, M.W. (2000) "The SUCCEED Longitudinal Database," Poster presentation at *SUCCEED 2000 Annual Meeting*, Greensboro, NC, March 28-29, 2000.

[10] Ohland, M.W. (1999) "Studying the Contribution of Programs at Eight Engineering Colleges toward Student Success," *Meeting of NSF Postdoctoral Fellows*, Washington, DC, September 16-17, 1999.

[11] Ohland, M.W., and T.J. Anderson (1999) "Studying the Contribution of Programs at Eight Engineering Colleges toward Student Success," *ASEE 1999*, Charlotte, NC, June 20-23, 1999.

[12] Neter, J., and M. H. Kutner, and C. J. Nachtscheim, and W. Wasserman (1996) *Applied Linear Statistical Models*, 4th edition, McGraw-Hill Companies, Inc., 1996.

[13] Ohland, M. W., G. Zhang, B. Thorndyke, and T. J. Anderson, "The Creation of the Multiple-Institution Database for Investigating Engineering Longitudinal Development (MIDFIELD)," accepted to *Proc. Amer. Eng. Ed.*, Salt Lake City, Utah, June 2004.

Author Information

GUILI 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.

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.

MATTHEW OHLAND is Assistant Professor in General Engineering at Clemson University. He has a Ph.D. in Civil Engineering with a minor in Education from the University of Florida, and served as the Assistant Director of the SUCCEED Coalition until 2000. His research is in freshman programs and educational assessment. He is the President of Tau Beta Pi, the engineering honor society, and will serve in that role until October 2006.

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.