International Journal of Engineering Education Vol. 31, No. 6(A), pp. 1448–1457, 2015
Printed in Great Britain
0949-149X/91 $3.00+0.00

Determining Graduation Rates in Engineering for Community College Transfer Students Using Data Mining*

MARCIA LAUGERMAN and DIANE ROVER
Department of Electrical and Computer Engineering, Iowa State University, Ames, IA 50011, USA.
E-mail:drmarcia@iastate.edu, drover@iastate.edu

MACK SHELLEY
Department of Political Science, Iowa State University, Ames, IA 50011-1204, USA. E-mail: mshelley@iastate.edu

STEVE MICKELSON
Department of Agricultural & Biosystems Engineering, Iowa State University, Ames, IA 50011, USA. E-mail: estaben@iastate.edu

This study presents a unique synthesized set of data for community college students entering the university with the intention of earning a degree in engineering. Several cohorts of longitudinal data were combined with transcript-level data from both the community college and the university to measure graduation rates in engineering. The emphasis of the study is to determine academic variables that had significant correlations with graduation in engineering, and levels of these academic variables. The article also examines the utility of data mining methods for understanding the academic variables related to achievement in science, technology, engineering, and mathematics. The practical purpose of each model is to develop a useful strategy for policy, based on success variables, that relates to the preparation and achievement of this important group of students as they move through the community college pathway.

Keywords: community college transfer; engineering; STEM policy; data mining; academic variables

1. Introduction

This study provides insight into community college (CC) transfer students who matriculate to the university in pursuit of an engineering degree. By synthesizing academic and demographic data from both institutions, we model graduation in engineering with academic and demographic variables from both institutions utilizing a data mining technique. Four cohorts totaling 472 CC transfer students were followed longitudinally for a minimum of six years to identify variables that were significantly correlated with graduation in engineering. The difficulty of obtaining, synthesizing, and analyzing trans-institutional data has resulted in few studies that address key retention variables and engineering persistence for the CC transfer student. The practical purpose of each model is to develop a useful strategy for policy, based on success variables, that relates to the preparation and achievement of this important group of students as they move through the CC pathway. The article also examines the utility of data mining methods for understanding the academic and policy variables related to achievement in science, technology, engineering, and mathematics (STEM).

The correlation between academic variables and graduation in engineering has been examined in numerous studies [1–9]. Of these studies only Tyson [1] included CC transfer students in the data. Persistence studies examine pre-college characteristics, which have been able to account for a small but meaningful percentage of variation in retention rates [8]. This study furthers the research by examining the role academic variables play in graduation, and extends it to include CC transfer students.

Internationally, CC-like institutions are increasingly recognized as 'have something very significant to offer to segments of the population'' [10]. ''There is a recognition around the world, and it manifests itself somewhat differently [in different countries], that community colleges, as one element of the higher education system, have something very significant to offer to segments of the population'' [11].

The study took into account in-state CC transfer students who were admitted to the College of Engineering (CoE) at a large Midwestern University from 2002–2005. It follows these transfer student cohorts longitudinally, each over a six-year period, to determine what academic integration characteristics contribute to their success in engineering using post-hoc graduation data. Strong prediction variables were discovered using the research strategy of boosted logistic regression to predict success in engineering for this group of CC transfer students. Boosted regression is used in this research to ascertain which academic prediction variables exert the most influence on the response

* Accepted 12 June 2015.

variable, which is graduation in engineering. The technique was developed in the artificial intelligence industry and is most frequently associated with data mining. The boosted regression logic is a relatively new strategy for retention and graduation rate research, but has shown success over traditional logistic regression models in prediction accuracy [12–15]. In addition to increased predictive accuracy, the results of boosted regression are intuitively easier to understand. It reports the percentage influence of each variable, instead of the odds ratios as reported in logistic regression or regression coefficients reported in traditional least squares regression to summarize the predictor variables' effects.

1.1 Background

In the United States, as many as 50 percent of college graduates turn to CCs for educational and professional advancement at some point in their educational careers [16–19]. Internationally, CCs educate 58 percent of students in Israel, 20 percent in Korea and France, and 26 percent in Japan. In addition, community college-like institutions are becoming increasing popular in about 30 different countries [10]. According to the American Association of Community Colleges (AACC), CCs provide a local, affordable, and low-risk path to develop and expand marketable skills [20]. The trend is especially strong for traditionally under-represented populations: women, minorities, rural students, veterans, and older Americans [16]. These groups are becoming increasingly central to the United States' mission to graduate more scientists and engineers [21]. However, many of these potential scientists and engineers leave this pathway before completing a four-year degree [22].

Research shows the importance of CC students, as they provide a rich source of engineers both in numbers and diversity. These engineers are vital for the jobs of the future. However, CC transfer students are difficult to analyze as a group because of their highly diverse nature. Furthermore, understanding and addressing persistence at the CC level is a multi-faceted task that must take into account fluctuating state funds and a diverse service population [23]. In addition, the enrollment patterns of CC students are complex and may involve multiple transfers across multiple institutions [24]. Admissions partnership programs improve the navigational success of CC transfer students to engineering [25].

1.2 Academic factors contributing to attrition in engineering

Among the external characteristics, the rigor of engineering curricula is cited as one of the most important variables contributing to student attrition, with calculus being the largest obstacle [8]. Students with a C average or less in calculus have a high probability of leaving engineering [5, 7]. Suresh [4] found that a majority of engineering majors earned a B minus or below, or withdrew from their first courses in Calculus and Physics while 20% of the students repeated these courses. Achievement in Calculus and Physics has been linked to engineering degree attainment [1, 3–4].

Most of the students who leave engineering do so before they have successfully completed these difficult courses [3]. Data show that students must acquire proficiency in these key foundational areas to succeed in engineering. In a longitudinal study of over 35,000 pre-engineering students at Purdue, 84% of those who leave engineering did so before they completed their pre-professional program [2]. Not all students who leave engineering do so because of low grades; many students leave engineering in good academic standing [8]. And not all students who stay in engineering have good grades. Retention has also been related to persistence and motivation [4], and to conscientiousness [26].

LeBold and Ward [27] found that the freshman year is critical to retention and that the best predictors of retention were the first and second semester grades and cumulative GPA. They found that students' perceptions of their problem-solving abilities in mathematics and science were also predictive of retention. Budny et al. [2] specifically looked at the effect of first-year course performance on graduation and found a strong correlation between first-semester GPA and graduation rates in engineering.

Other researchers have also found that the single fundamental variable predicting retention in STEM fields is grade point average [6]. Whalen and Shelley [6] found a dramatic increase in six-year retention and graduation rates for as little as a 0.10 increase in GPA for STEM majors. Earlier research by Strenta, Elliot, Adair, Matier, and Scott [28] found that low grades were the most common predictor for all students leaving science and engineering courses. Schools have found that success strategies such as tutoring, supplemental instruction, and counseling are effective in helping students complete these high-risk courses [2, 29].

Pre-college characteristics account for a small but meaningful percentage of variation in retention rates [8]. However, research shows that pre-engineering success measures are weaker predictors of retention in engineering than are grades in core engineering courses [2–3]. Further, the combination of first-year course grades is a stronger predictor of success than the grade in any single course.

Various data analysis methods have been applied to predict retention and graduation rates by using academic and demographic variables. Conventional predictive models have used logistic regression. Other data analysis methods existing in the literature are summarized by Li, Swaminathan, and Tang [30]:

- Stepwise/Hierarchical Multiple Regression
- Longitudinal Data Analysis
- Covariate Adjustment
- Two-Step Design
- Exploratory Factor Analysis
- Structural Equation Modeling
- Discriminant Analysis
- Classification Tree

This research uses longitudinal data analysis, with covariate adjustment, utilizing boosted logistic regression to determine which of the included academic variables exert the greatest percentage influence toward graduation in engineering for these cohorts of CC transfer students. This approach is designed to discover which predictors are associated with the highest level of influence on academic achievement for success of this important group of students. This strategy thus provides a roadmap for CC student success in transitioning into engineering majors at four-year institutions of higher education.

1.3 Objectives

The objectives of this study are to:

1. Develop an overall boosted logistic regression model using academic and demographic variables that is predictive of graduation in engineering for CC transfer students who have completed the Basic Program [BP] in engineering, as described below.
2. Determine the fit statistics for this model by comparing to actual graduation rates. This result determines the utility of this data mining method for understanding the academic and policy variables that are most closely related to achievement in engineering.
3. Report on the levels of student achievement for academic variables that maximize their chances for success in engineering. Doing so will address the preparation and achievement of this important group of students as they move through the CC pathway toward an undergraduate degree in engineering.

2. Methods

Using the University's institutional research data, 11,632 records were obtained for students who were admitted to the CoE in fall semesters from 2000 to 2010 (inclusive). Two groups of students were investigated based on their admission status to the University: 10,441 who were admitted directly from high school (DHS), and 1,191 who transferred from in-state CCs. A subset of this group comprises 472 CC students who were admitted to the CoE between 2002 and 2005 (inclusive). This group was selected to provide sufficient time for graduation in engineering. *T*-tests were administered between fall semester entries and spring semester entries, and found no differences in demographic or academic variables. From these larger datasets two groups of students were investigated based on their admission status to the University: those who were admitted directly from high school (DHS), and those who transferred from in-state CCs.

2.1 Variables

In this study core-course offerings (called the Basic Program [BP] in engineering) are examined in detail since they have been shown to have the most predictive accuracy in relevant research [1–3]. The BP is a common set of courses required of all engineering students at the University. All students must successfully complete the BP with a minimum C average (2.0 on a 4.0 scale) to graduate in engineering. This program consists of two semesters of calculus, one semester of chemistry, one semester of physics, two semesters of English, and one semester of engineering fundamentals with computer programming. The academic variables are presented that exert the most influence on graduation in engineering,

The academic variables included in the study that were transferred from the CC included: individual BP course grades, overall GPA for all transferred BP courses, overall transfer GPA, number of credits transferred toward the BP, and the total number of credits transferred. The academic variables obtained from the University after transfer were: individual BP course grades; overall GPA for all BP courses taken at the University; the first fall, first spring, and first-year GPA at the University; and the number of credits the first fall, first spring, and first year at the University. Since a CC student has the option of transferring some or all of the BP courses, students' BP course grades are included from both the CC and the University.

The student background variables included are: gender, ethnicity, and learning community participation. Other typical demographic variables have too many missing values in the dataset to be included in the study. It is assumed that the academic and background variables for the groups of fall cohorts entering engineering from 2002–2005 represent random, independent, normally distributed samples. The sample sizes and the Central Limit Theorem help to validate the normality assumption

[31]. Density function graphs were examined for each high-effect independent variable, with no major departures from normality observed except for a slight left skew, which is expected in GPA measures.

CC grades can provide a missing piece of the puzzle in graduation and retention research [1]. Introducing CC course grades increases variability within the dataset, so results that include grades from CC courses are separated from results that include grades in courses taken at the University. It is assumed that the groups of CC students taking the courses at either institution are equivalent. No statistical information was found contrary to this assumption.

2.2 Students

The demographic characteristics of the 11,632 students who enrolled in the CoE from fall 2000 to fall 2010 were as follows:

- Female: 6.8%
- Black: 3.5%
- White: 84.5%
- Hispanic: 1.6%
- American Indian: 0.9%
- Asian: 3.8%
- US Citizen: 92.9%

Separating the data into DHS and CC admits, the distribution of the CC demographic data was compared with that of DHS admits over the same time period. Using a Pearson chi-square analysis, there were no differences in the demographic characteristics ($p > 0.10$), except for females. The proportion of female students was significantly less ($p < 0.0001$) for CC admits than for DHS admits to engineering. It is assumed that any sub-group of these students will have similar characteristics. The sample is large enough for the observations to yield sufficient power for the statistical tests to be valid [31].

Table 1 shows the background characteristics by admission status to the CoE from fall 2000 to fall 2010. It compares background characteristics for CC transfer admits with students admitted directly from high school (DHS) over the same time period. This table must be interpreted with caution, since the data include background characteristics for only 50% to 70% of the CC transfer students. Even considering this lack of complete data, it appears that this group of CC transfers come in with weaker academic backgrounds as measured by mathematics American College Test (ACT) scores (or equivalent mathematics Scholastic Aptitute Test (SAT) scores) and high school GPAs.

Table 2 presents GPA data for CC transfer students, compared to DHS students at significant intervals (end of the first fall, and end of the first year at the University), as well as the University portion of the BP GPA. These show statistically significantly lower GPAs for CC transfer students at each interval and for the courses taken at the University toward the BP. Other research agrees with this finding. Tsapogas [32, p. 6] notes that GPAs tend to be lower for transfer students: "Science and engineering graduates with lower undergraduate grade point averages are more likely to have attended community college than are graduates with higher grade point averages". These lower GPAs may lead to lower grades in the engineering BP and lower retention and graduation rates.

Table 1. Background Characteristics by Admission Status

Admit Type	**Math ACT**	***n***	**HS GPA**	***n***
DHS	28.0	9,849	3.63	10,441
CC Transfer	25.0*	585	3.24*	585

*$p < 0.05$.

Notes: Admitted to the CoE during summer and fall Semesters 1999–2009.

Measured as enrolled in engineering as of fall semester for the years 2000–2010.

ACT = American College Testing, HS = High School, GPA = Grade Point Average, *n* = sample size, DHS = Direct from High School, CC = Community College.

Table 2. Admit Status and GPA

Admit Status	**DHS (n = 9,065)**	**CC Transfer (n = 1,011)**
First fall GPA	2.72	2.31*
First year GPA	2.78	2.42*
University BP GPA	2.71	2.32*

*$p < 0.05$.

Notes: Admitted to the CoE during summer and fall semesters 1999–2009.

Measured as enrolled in engineering as of fall semester for the years 2000–2010.

DHS = Direct from High School, CC = Community College, BP = Basic Program in Engineering, GPA = Grade Point Average.

2.3 Analysis method

This research uses the Stata data analysis package with the AdaBoost feature. Stata is a general-purpose statistical software package created in 1985 by StataCorp. The "boost" command within Stata starts the boosting algorithm described in Hastie, Tibshirani, and Friedman [13] to develop models that predict graduation in engineering. The overall model shows the academic variables having the highest influence on graduating in engineering for this group of CC transfer students. Strengths of the boosting algorithm include that interactions and nonlinearities need not be explicitly specified and that categorical variables do not need to be transformed [13]. Boosted logistic regression is used to determine which of the included academic variables exert the greatest influence on predicting graduation in engineering, while controlling for background student characteristics variables. Boosting works by sequentially applying a classification algorithm to reweighted versions of the training data and then taking a weighted majority vote of the sequence of classifiers thus produced. For many classification algorithms, this simple strategy results in dramatic improvements in performance. This approach can be understood in terms of well-known statistical principles, namely additive modeling and maximum likelihood. For the two-class problem, boosting can be viewed as an approximation to additive modeling on the logistic scale using maximum Bernoulli likelihood as a criterion. This is not a black box or a capitalization on chance methodology, in that all the academic and demographic prediction variables have a basis in theory.

The AdaBoost procedure trains the classifiers on weighted versions of the training sample, giving higher weight to cases that are currently misclassified. This is done for a sequence of weighted samples, and then the final classifier is defined to be a linear combination of the classifiers from each stage [12–15]. However, missing values do create problems for boosted regression and must be dropped from the analysis. In this analysis, less than 9% of the dataset was dropped; therefore we did not employ any of the missing data analysis subroutines that were available.

With the boosted regression technique, correlated predictors can be incorporated into the model, such as using first-semester GPA, second-semester GPA, and first-year GPA as concurrent predictors. The mean-square error (MSE) term incorporates the error for each exogenous variable, including correlated variables, thus taking into account the additional error from correlated terms. Also, the separation of training data and test data helps guard against over-fitting that may arise in the context of correlated data. All of the variables in the final models are tested for collinearity using the variance inflation factor (VIF). Generally, VIF statistics less than 5 are considered acceptable [33]. Therefore, no highly correlated variables are included in the final models.

The boost command determines the number of iterations that maximize the likelihood, or, equivalently, the pseudo R^2 values. Pseudo R^2 values are computed for both the training and the test data within the model. The training model contains 80% of the dataset and the test model contains the other 20% of the dataset. These percentages were varied to see the effect on the pseudo R^2 values. No statistical reason was found to change these percentages.

Pseudo R^2 is a measure of predictive accuracy, not model fit; it can be small for a properly specified model and can be large even when the model is misspecified. The pseudo R^2 values illustrate how much of the variation in graduation rates is explained by variation of the prediction variables in the model. The pseudo R^2 statistic is defined as 1–L1/L0, where L1 and L0 are the log likelihood of the full model and intercept-only model, respectively. Unlike the coefficient of determination, R^2, value given in least squares regression, the pseudo R^2 value is an out-of-sample statistic (a smaller percentage of the population, generally 20%). Out-of-sample R^2 values tend to be lower than in-sample R^2, which is the case in this study. The reason 1-L1/L0 is called pseudo R^2 is that its formula resembles the coefficient of determination, R^2, which is equal to 1- SSE/SST, where SSE is the sum of the squares due to error (unexplained variation) and SST is the total sum of squares (explained plus unexplained variation). Larger R^2 (or pseudo R^2) values indicate better fit of the model, meaning the amount of unexplained error is small. For that to happen, the ratio L1/L0 needs to be small, which means L1 needs to be much smaller than L0. This implies that the full model is better than the null model (similar to having a model with small SSE) [33].

Once the models are determined, the model-fit statistics are analyzed using pseudo R^2 values of the training and test data, the MSE values, and the root mean square error values (RMSE). MSE values show the amount of variation in the chi-square goodness of fit test statistic that is accounted for in the model and RMSE values determine the extent to which the estimated model differs from the actual on average.

Graduation rates tables are created to compare the predicted and actual graduation rates with levels of achievement for the highest-effect variables. The idea is to create recommended thresholds of achievement based on this group of CC transfer students. There was a naturally occurring break in the graduation rates above 40% and again above 70%. The levels of achievement for the highest-effect variables are measured at these graduation rates, resulting in recommended thresholds of achievement.

3. Results

For the period fall 2002 to fall 2005 (inclusive) there were 472 in-state CC transfers to the CoE. The average graduation rate for this group of students was 54%. The characteristics of in-state CC transfer students were measured in another study [34] based on a slightly different subset of the data, but their characteristics are assumed to be similar to the 2002 to 2005 admit group. Of the characteristics reported in Laugerman and Mickelson [34]:

- 37% graduate in Mechanical Engineering
- most take jobs in Iowa (76% or more)
- 98% are residents of Iowa
- the average student transfers 60 credits
- the average transfer GPA is 3.2
- the average ISU GPA at Graduation is 2.9
- 3.4% are female
- 8.3 % are non-white US citizens
- the average age of the graduate is only slightly older than the average DHS graduate, and
- the average time to graduation is about 7 semesters after admission

3.1 Overall model

The overall model is determined at the point when a student has completed the BP courses, which may occur later than after the first year at the University. The four highest-influence variables in the overall model (Table 3) are the same as for the one-year model (not specified), showing the sustained importance of these variables in predicting graduation in engineering. Table 3 shows that the first-year GPA exerts 39.5% of the influence on graduation in engineering, while the total CC BP credit hours transferred exerts 22.0% influence on graduation in engineering.

Table 4 compares the boosted model predictions in 0.20 increments to the actual graduation rates in engineering, including the levels of high-influence variables. This table illustrates how the predicted probabilities compare to the actual rates of earning an engineering degree. It shows how the model over-predicts graduation rates at lower levels. Particularly noteworthy are the small differences in parameter values between the 77% and the 98% actual graduation rates, suggesting that there is a big difference in graduation rates even for small increases in high-influence variables. For the highest-influence variable, this would mean the difference between a 2.74 and a 3.09 university first-year GPA.

Table 5 reduces the graduation rates from Table 4 into three naturally occurring categories. This table is useful for recommending thresholds of achievement of high-effect variables, particularly at the completion of the BP in engineering. In order of highest-effect variables, a CC transfer student should strive to achieve a University first-year GPA of 3.04 or above and transfer at least 19.3 credits toward BP courses. For this group of students, these traits resulted in a 94% or better probability of graduating in engineering, which is a significant improvement over the average graduation rate of 54%. The table also shows how the model may be under-predicting the graduation rates at higher levels.

Table 3. Overall model: Variable influence factors for highest-effect variables

Academic Variable	Influence on Earning an Engineering Degree
First Year GPA	39.5%
CC BP transfer credit hours	22.0%
First fall credits completed at the university	7.2%
First fall GPA at the university	6.0%
CC BP transfer GPA	5.4%
First year credits completed at the university	4.0%
Credit in Physics I completed at the university	3.4%
Credit in Calculus I completed at the university	1.0%

Notes: For Fall 2002 to Fall 2005 CC transfer admissions. The total percentage influence is 100%—some low percentage variables are omitted. DHS = Direct from High School, CC = Community College, BP = Basic Program in Engineering, GPA = Grade Point Average.

Table 4. Overall model comparison of graduation probabilities by highest-effect parameters in 20% increments

Predicted Probability of Earning Engineering Degree	Actual Probability of Earning Engineering Degree	CC BP Transfer GPA	First Fall University GPA	First Year University GPA	First Fall University Credit Hours Completed	Number of BP Transfer Credits	*n*
10%–20%	2%	2.82	1.54	1.56	11.2	10.5	112
20%–40%	6%	2.95	1.95	2.16	11.3	16.3	48
40%–60%	46%	3.01	2.38	2.38	12.3	15.7	41
60%–80%	77%	3.03	2.49	2.74	12.4	16.9	57
80%–100%	98%	3.39	3.01	3.09	12.8	20.1	160
Average	54%	3.10	2.36	2.45	12.5	16.2	418

Notes: DHS = Direct from High School, CC = Community College, BP = Basic Program in Engineering, GPA = Grade Point Average.

Table 5. Overall model comparison of graduation probabilities by highest-effect parameters in 30% increments

Predicted Probability of Earning Engineering Degree	Actual Probability of Earning Engineering Degree	CC BP Transfer GPA	First Fall University GPA	First Year University GPA	First Fall University Credit Hours Completed	Number of BP Transfer Credits	*n*
10%–40%	3%	2.86	1.66	1.74	11.2	12.2	160
40%–70%	53%	2.95	2.34	2.42	12.5	16.6	58
70%–100%	94%	3.33	2.93	3.04	13.5	19.3	41
Average	54%	3.10	2.36	2.45	12.5	16.2	418

Notes: CC = Community College, BP = Basic Program, GPA = Grade Point Average.

Table 6. Variance inflation factor values for overall model

Variable	Variance Inflation Factor
First fall university GPA	3.47
First year university GPA	3.47
Number of BP transfer credits	2.60
CC BP transfer GPA	1.52
First fall university credit hours completed	1.48

Notes: A value less than 5 indicates low or no collinearity (Levine, 2008).
CC = Community College, BP = Basic Program, GPA = Grade Point Average.

Table 6 shows the variance inflation factor (VIF) values for all of the variables in the overall model. Based on the VIF results the variables have minimal collinearity (i.e., redundancy).

3.2 Model fit statistics

The model fit statistics for the overall model are provided in Table 7. In the table, the test R^2 is the amount of variation in the graduation rates that is explained by the variables used to test each model. In the overall model 35.4% of the variation in the graduation rate is explained by variation of the parameters in the model.

Table 7. Model fit statistics

Test R^2	Training R^2	MSE	RMSE
0.354	0.901	0.080	0.282

Notes: MSE = mean square error, RMSE = root mean square error.

The training R^2 is the amount of variation in the graduation rate that is explained by the variables used to create (train) the model. This is expected to be much higher than the test rates since 80% of the observations are used to create the model. In the overall model the training R^2 value is 0.901 (90.1%). The MSE of 0.080 shows the amount of variation in the chi-square goodness of fit test statistic that is accounted for in the model and the RMSE of 0.282 determines the extent to which the estimated model differs from the actual on average.

3.3 Summary

For this group of CC transfers to the CoE, Table 8 summarizes the variables that exerted the highest-influence on graduation in engineering and the recommended thresholds of achievement for these variables. The graduation rate in engineering for students achieving the recommended levels of these high-influence variables was 94%. Note the number of CC BP credits recommended was 19.3, which is a majority of the approximately 27 credits included in this program.

4. Discussion

Using research from a United States CC and University, strong prediction variables were discovered using the research strategy of boosted logistic regression to predict success in engineering for this group of CC transfer students. In the analysis, the most influential predictors of graduation in engineering for this group of CC transfer students were the university GPA after transfer and the number of

Table 8. Summary of influence variables and recommended thresholds

Influence variable	% Influence on earned engineering degree	Recommended Threshold
First-year University GPA	39.5%	3.04
CC BP transfer credit hours	22.0%	19.3
First fall credits completed	7.2%	13.5
First fall GPA	6.0%	2.93
CC BP transfer GPA	5.4%	3.33

Notes: CC = Community College, BP = Basic Program, GPA = Grade Point Average.

credits transferred toward BP courses. The results suggest that there is a big difference in graduation rates even for small increases in high-influence variables. For the highest-influence variable, this would mean the difference between a 2.74 and a 3.09 university first-year GPA. Transferring more credits toward BP courses was another result of the study. Both could be measures of preparation and persistence of the CC transfer student. Since the progression toward an engineering degree begins at Calculus I, students who are calculus-ready are better prepared to study engineering than are those who start in remedial mathematics course work. Furthermore, the number of BP credits measures persistence in Calculus, Physics, and Chemistry, all high predictors of success in engineering.

4.1 Limitations

This study uses information from multiple CCs, but is limited to a single state and one university. Although previous research supports the conclusions in this article, these results may not be representative of other state or national CCs or universities. This study includes students who dropped out or stopped out and returned to the University, but does not track the students who left and did not return in the six-year time period. Some of these students undoubtedly were successful in obtaining a certificate or degree from another institution, but these students were not included in the study. Students who did not start in engineering but later changed majors to the CoE were not included because of the small number of students involved and the complication these data would have added to the research. In addition, this research does not include information about credentials earned at the CC, such as associate's degrees or other certificates, to be able to focus on just the BP course information.

Model fit statistics are always important in determining the success of predictive models. Fitting models that predict graduation in engineering is sufficiently complex that it is unrealistic to expect any model to explain a very high proportion of the variation in student success. The most easily understood model fit statistic is the test pseudo R^2 value, which measures the amount of variation in the graduation rates that is explained by the variables in the model. The overall model explains about 35% of the variation in graduation rates in engineering, with a parsimonious number of academic variables. This is a high rate for a predictive model [8]. However, the overall model is measured at the point where a student has completed the BP courses, and most of the attrition in engineering already may have occurred by that point. On average, the model tends to over-predict graduation rates at lower levels, and under-predict graduation rates at higher levels. Other problems with the model fit can be explained by a number of circumstances:

1. Missing variables. Social and financial constructs from the models such as cognitive reasoning ability and quantitative reasoning ability are missing.
2. Measurement error of the variables included in the model.
3. Specification error of the variables. Although nonlinearities of exogenous variables need not be explicitly explained in boosted logistic regression models, interactions between variables and transformations of the endogenous variable are not examined in this work.

In addition, the explained variation in the models does not necessarily imply casualty. Instead, the models can only imply correlations between the exogenous variables and the response variable. Even so, other research studies support the ability of the academic variables to predict graduation rates in engineering [1–3].

5. Conclusions

This study provides information about academic pathways for CC transfer students into engineering by identifying academic variables and levels of these academic variables that highly influence success rates for this important group of students. Results from this study are informative to a global audience interested in the opportunities offered by CC-like organizations interested in preparing students to succeed in engineering. Based on the variables used in the study, the two most influential predictors of graduation in engineering are consistently the overall GPA at the University (after transfer) and the number of CC credits transferred that apply to the BP (core courses) in engineering. Even very small increases in GPA have significant effects on increasing the graduation rates in engineering. In addition, students who transfer more credit toward completing the BP in engineering have higher graduation rates.

Although this study does not consider graduation in a major other than engineering, many of the students who leave engineering do graduate successfully from the university, which makes for a logical extension of this research. In addition, this study does not have the power of a meta-analysis, which would validate and extend the research findings. To test these findings further, the models could be tested against other cohorts of CC transfer students who have had time to complete a degree in engineering. Future research could use this information to develop a classification system to predict success in engineering. Qualitative research that examines how to raise levels of academic variables also would be valuable.

To the degree that academic strategies are able to predict success in engineering the levels of achievement in key academic variables are useful. They can be used to design the best course of research and utilize programs for skills improvement, especially in mathematics and science, as needed. This will help illuminate a successful pathway to an engineering degree for a CC student, and may be able to help increase the number of students who successfully navigate this pathway.

Acknowledgements—This material is based upon work supported by the National Science Foundation under Grant No. 0653236. Any opinions, findings, and conclusions or recommendations expressed in this material are those of the author(s) and do not necessarily reflect the views of the National Science Foundation. We also acknowledge the cooperation and support of Kari Henson and other faculty, staff, and administrators at Des Moines Area Community College as well as Jason Pontius at Iowa State University for research support.

References

1. W. Tyson, Modeling engineering degree attainment using high school and college physics and calculus coursetaking and achievement, *Journal of Engineering Education*, **100**(4), 2011, pp. 1–18.
2. D. Budny, W. LeBold and G. Bjedov, Assessment of the impact of freshman engineering courses, *Journal of Engineering Education*, **87**(4), 1998, pp. 405–411.
3. J. Levin and J. Wyckoff, *Identification of student characteristics that predict persistence and success in an engineering college at the end of the sophomore year: Informing the practice of academic advising*. Division of Undergraduate Studies Report No. 1990, **1**, Pennsylvania State University.
4. R. Suresh, The relationship between barrier courses and persistence in engineering, *Journal of College Student Retention*, **8**(2), 2006, pp. 215–239.
5. C. Veenstra, E. Dey and G. Herrin, A model for freshman engineering retention, *Advances in Engineering Education*, **1**(3), 2009, pp. 1–33.
6. D. Whalen and M. Shelley, Academic success for STEM and non-STEM majors, *Journal of STEM Education*, **11**(1 and 2), 2010, pp. 45–60.
7. G. Zhang, Y. Min, M. Ohland and T. Anderson, The role of academic performance in engineering attrition. In *Proceedings of the 2006 American Society for Engineering Education Conference and Exposition*, Chicago, IL.
8. G. Zhang, T. Anderson, M. Ohland, R. Carter and B. Thorndyke, Identifying factors influencing engineering student graduation and retention: A longitudinal and cross institutional study, *Journal of Engineering Education*, **93(**4), 2004, pp. 313–320.
9. S. Palmer, Modeling Engineering Student Academic Performance Using Academic Analytics, *International Journal of Engineering Education*, **29**(1), 2013, pp. 132–138.
10. E. Redden, News: The 'Community College' Internationally—Inside Higher Ed. *Home—Inside Higher Ed*. N.p., 16 June 2010. Web. 7 June 2015.
11. T. Cornick, as cited by E. Redden, News: The 'Community College' Internationally—Inside Higher Ed. *Home—Inside Higher Ed*. N.p., 16 June 2010. Web. 7 June 2015.
12. J. Friedman, T. Hastie and R. Tibshirani, Additive logistic regression: A statistical view of Boosting, *The Annals of Statistics*, **28**(2), 2000, pp. 337–407.
13. T. Hastie, R. Tibshirani and J. Friedman, *The elements of statistical learning: Data mining, inference, and prediction.* New York: Springer-Verlag, 2001.
14. J. Long and J. Freese, Regression Models for Categorical Dependent Variables Using Stata, Revised Edition. College Station, TX: Stata Press, 2003.
15. M. Schonlau, Boosted regression (boosting): An introductory tutorial and a Stata plugin, *The Stata Journal*, **5**(3), 2005, pp. 330–354.
16. American Association of Community Colleges, *July 2011 report Improving student transfer from community colleges to four-year institutions*. Retrieved from http://advocacy.collegeboard.org/sites/default/files/11b3193transpartweb110712.pdf, accessed 1 April, 2014.
17. D. Baime, Obama's best bet: $5 billion for community colleges, *The Chronicle of Higher Education*, **58**(3), 2011, A30.
18. S. Handel and R. Williams, The promise of the transfer pathway, The Initiative on Transfer Policy and Practice, College Board, 2012.
19. C. Mullin, *The road ahead: A look at trends in the educational attainment of community college students*. (Policy Brief 2011-04PBL). Washington, DC: American Association of Community Colleges.
20. American Association of Community Colleges. *AACC statement on trends in college pricing and aid 2009*. Retrieved from http://www.aacc.nche.edu/newsevents/News/articles/Pages/102020091.aspx, accessed 1 July 2013.
21. National Science Board. *National science board science and engineering indicators.* 2010 (Appendix Table 2–35). Washington, DC: National Science Board. Retrieved from http://www.nsf.gov/statistics/seind10/append/c2/at02-35.pdf, accessed 1 July 2013.
22. National Research Council. *Enhancing the community college pathway to engineering careers.* Washington, DC: The National Academies Press, 2005.
23. T. Bailey and M. Alfonso, *Paths to persistence: An analysis of research on program effectiveness at community colleges*. Indianapolis IN: Lumina Foundation for Education, 2005
24. A. McCormick, Swirling and double-dipping: New patterns of student attendance and their implications for higher education. *New Directions for Higher Education* (121), 2003, pp. 13–24.
25. M. Laugerman, M. Shelley, S. Mickelson and D. Rover, An initial evaluation of a STEP initiative: The Engineering Admissions Partnership Program designed to increase success of community college transfer students, *International Journal of Engineering Education*, **29**(5), 1260–1269.
26. C. Hall, P. Kauffmann, K. Wuensch, W. Swart, K. Deurquidi, O. Griffin and C. Duncan, Aptitude and personality traits in retention of engineering students, *Journal of Engineering Education*, **104**(2), 167–188.
27. W. LeBold and S. Ward, *Engineering retention: National and institutional perspectives*. Proceedings, 1988 ASEE Annual Conference. ASEE, 1988, pp. 843–851.
28. C. Strenta, R. Elliot, R. Adair, M. Matier and J. Scott, Choosing and leaving science in highly selective institutions, *Research in Higher Education*, **35**(5), 1994, pp. 513–547
29. M. Shelley and K. Hensen, The Impact of supplemental instruction: Results from a large, public, Midwestern university. *Journal of College Student Development*, **44**(2), 2003, pp. 250–259.
30. Q. Li, H. Swaminathan and J. Tang, Development of a classification system for engineering student characteristics affecting college enrollment and retention, *Journal of Engineering Education*, **98**(4), 2009, pp. 361–376.
31. D. Levine, *Statistics for managers using Microsoft Excel* (5th ed.). Upper Saddle River, NJ: Pearson Prentice Hall, 2008
32. J. Tsapogas, *The Role of community colleges in the education of recent science and engineering graduates.* (NSF 04-315). Arlington, VA: National Science Foundation, 2004
33. A. Agresti and B. Finley, *Statistical methods for the social sciences* (4th ed.). Upper Saddle River, NJ: Pearson Prentice Hall, 2009
34. M. Laugerman and S. Mickelson, *Characteristics of Community College Transfer Students that Successfully Matriculate and Graduate in Engineering*, American Society of Engineering Education Annual Meeting, Vancouver, British Columbia, Canada, June 2011.

Marcia R. Laugerman is a licensed professional Industrial Engineer with the BS and MS degrees in Industrial Engineering and the PhD in Industrial Technology. She is currently a Research Fellow with Iowa State University, Department of Electrical and Computer Engineering. She has extensive experience as an educator and advisor in STEM higher education fields and a developing research portfolio in K-20 STEM. Dr. Laugerman's teaching expertise is in engineering fundamentals, engineering economics, statistics, and computer information systems. She has experience in designing collaborative, cross-disciplinary courses; implementing learning management systems; utilizing constructivist course designs; and structuring asynchronous courses. Her research projects have been funded by the National Science Foundation and the United States Department of Education-Institute of Education Sciences.

Diane T. Rover is a University Professor in the Department of Electrical and Computer Engineering at Iowa State University. Dr. Rover was Associate Dean for Academic and Student Affairs in the College of Engineering from 2004–2010. Prior to that, she served as associate chair for undergraduate education in the Department of Electrical and Computer Engineering from 2003–2004. She began her academic career at Michigan State University. She received the B.S. in computer science in 1984, and the M.S. and Ph.D. in computer engineering in 1986 and 1989, respectively, from Iowa State University. She has served on the Engineering Accreditation Commission of ABET since 2009. From 2006–2009, she served on the IEEE Committee on Engineering Accreditation Activities. Since 2009, she has held the positions of secretary/ treasurer, program chair, chair-elect, chair, and past chair of the ASEE ECE Division. From 2000–2008, she led the Academic Bookshelf column as a senior associate editor for the *ASEE Journal of Engineering Education*. Her teaching and research has focused on embedded computer systems, reconfigurable hardware, integrated program development and performance environments for parallel and distributed systems, visualization, performance monitoring and evaluation, and engineering education. Dr. Rover is a 2012 ASEE Fellow and member of the IEEE Computer Society, the IEEE Education Society, and ASEE.

Mack Shelley is a University Professor at Iowa State University with joint appointment in the Departments of Statistics and Political Science. From 2003–2007 he served as Director of the Research Institute for Studies in Education (where he also was Coordinator of Research from 1999–2003), and from 1999–2007 was a Professor in the Department of Educational Leadership and Policy Studies. His research, external funding, and teaching focus on applications of statistical methods to public policy and program evaluation, with emphasis on education policy and programs. He has received funding from numerous federal agencies, state agencies, and other organizations. He serves regularly as a statistical consultant for researchers, administrators, program staff, and students, and has received awards for research, teaching, and professional practice. He has served as co-editor and a member of the editorial boards of a number of scholarly journals. He currently serves as Chair of the Department of Political Science.

Steven K. Mickelson is a professor in the Department of Agricultural and Biosystems Engineering (ABE) at Iowa State University. He serves as the Chair for ABE and is the inaugural recipient of the Charles R. and Jane F. Olsen Professorship in Engineering. The named professorship was made possible through the generosity of Charles and Jane Olsen of Leawood, KS to recognize an outstanding faculty member with research and teaching emphasis in the areas of sustainability, clean water and the environment. Dr. Mickelson's teaching expertise is in engineering computer graphics, engineering design, and engineering workplace competency development and his research focus areas includes evaluation of agricultural best practices for reducing surface water contamination and the scholarship of teaching and learning. Prior to becoming Chair of ABE, he was Director of the Center for Excellence in Learning and Teaching and the Co-director of Learning Communities at ISU for three years. He received his agricultural engineering degrees from Iowa State University in 1982, 1984, and 1991, respectively.