AC 2007-1592: PREDICTORS OF SUCCESS IN THE FIRST TWO YEARS: A TOOL FOR RETENTION

Paul Kauffmann, East Carolina University

Paul J. Kauffmann is Professor and Chair in the Department of Engineering at East Carolina Univerisity. His industry career included positions as Plant Manager and Engineering Director. Dr. Kauffmann received a BS degree in Electrical Engineering and MENG in Mechanical Engineering from Virginia Tech. He received his Ph.D. in Industrial Engineering from Penn State and is a registered Professional Engineer.

Tarek Abdel-Salam, East Carolina University

TAREK ABDEL-SALAM is an Assistant Professor in the Department of Engineering at East Carolina University. Dr. Abdel-Salam received a Ph.D. in Mechanical from Old Dominion University. His research interests include educational effectiveness in engineering education, energy management, and thermal / fluid systems.

John Dail Garner, East Carolina University

John Dail Garner is a Lecturer in the Department of Engineering at East Carolina University. Mr. Garner received an MS in Mechanical Engineering from North Carolina State University. His research interests include recruitment and retention in engineering education, as well as energy production and conservation in agriculture.

PREDICTORS OF SUCCESS IN THE FIRST TWO YEARS: A TOOL FOR RETENTION

Abstract

Retention is a significant issue in engineering education. The ability to identify factors in student records which best predict academic success can be a very important tool in developing and implementing the timely and focused interventions which are an essential part of a strategic plan to improve retention rates. This paper presents a study conducted to improve retention rates by using step wise regression to identify the most significant factors to predict undergraduate grade point average at the end of the freshman and sophomore years. The model examines standardized test scores, rank in high school class, and various measures of high school grade point average for three different years of performance. The results show that, for this sample of first and second year students, un weighted high school grade point average and rank in high school graduating class are the most important predictors of college grade point average success. Standardized test scores were not significant predictors.

Introduction

Retention of engineering students is a continuing concern among university academic programs nationwide. In improving retention, engineering educators have spent significant effort in identifying relationships between various measures of success and prediction variables. In this way, it may be possible to identify targeted interventions to improve success or prevent failure. As a result of these efforts, a variety of multi-variable models have been developed to predict various measures of student success using a range factors.

In one example, Takahira et al.[1] found that the primary factors associated with persistence in an engineering statics course were GPA and SAT-math scores. Another study reported a positive effect of an entrepreneurship program on GPA and retention.[2] Other researchers found scores from a non-technical, writing assignment was a predictor of academic success of freshmen engineering students as measured by cumulative grade point average after completion of the first two semesters.[3]

Other models have been more complex. Student success and persistence were examined by French et al.[4] using hierarchical linear regression. They examined both quantitative variables (SAT scores, high school rank, university cumulative grade point average) and qualitative variables (such as academic motivation and institutional integration). For measures of success they used junior and senior GPA, university enrollment and major enrollment over six and eight semesters. The study found that SAT scores, high school rank, and gender were significant predictors of GPA and that an orientation course was ineffective.

Zhang et al.[5] evaluated a number of factors and their impact on engineering student success as measured by graduation rate. Using a multiple logistic regression model and data from nine institutions, they examined the impact on college graduation of high school GPA, gender, ethnicity, quantitative and verbal SAT scores, and citizenship and their impact on graduation.

The study reports that, high school GPA and quantitative SAT score were the only significant factors for all models tested and the significance of other predictors varied among institutions.

Another study[6] examined two frequently used selection variables, high school GPA and ACT composite test score, as predictors of college graduation success in a minority engineering program. It found that, the only significant predictor of engineering graduation was high school GPA. Both high school GPA and ACT scores were predictors of college GPA but only high school GPA was a predictor of graduation in engineering.

The objective of this study is to complement this literature by more closely examining performance contributors which impact success in the first two years of engineering study. Generally, student success or failure, as measured by GPA in this time interval, results in changes in major or withdrawal from the university; the actions which result in retention failures for the engineering degree. Consequently, the primary focus of this study is to determine which variables can predict freshmen and sophomore GPA. With this information, program faculty can identify appropriate interventions to improve retention. The next section provides background information about the university and the engineering program which conducted this study.

Background Information

Over the last decade, there has been a major shift in the economic base of eastern North Carolina. For generations, the regional economy was driven by agriculture and in particular by textile and tobacco production. In recent years, this economy has made a major transition to a manufacturing, military, and government / service based economy. The region's ability to maintain this momentum and continue to grow technology driven businesses is, in large part, dependent on attracting and retaining engineering expertise. However, local, national and global firms often have difficulty attracting and retaining engineers in a region that is primarily comprised of rural towns and small cities. The ability to grow local engineering and technology talent, with family roots in the region, was identified as an important element in addressing this issue.

With a student enrollment in excess of 23,000, East Carolina University is the third largest institution in the sixteen member University of North Carolina System. ECU has been an engine of development and progress in the region, primarily due to a medical school and the leading College of Education in North Carolina. To further enhance the university's capability, a general engineering program was approved and accepted its first students in fall 2004 into a systems engineering concentration. Engineering management and bio process engineering were added in 2005.

Many students in the program come from small, rural communities and are often first generation college students. Consequently, retention was expected to be a significant challenge and student performance has been closely tracked relative to data gathering and analysis.

Purpose of the Study

An essential goal for the new engineering program was to create a collaborative learning environment which would promote retention and graduation rates which outperform national standards. To support efforts to meet this program goal, a primary focus of program analysis has been to identify factors (student characteristics) which are indicators of future academic success or, conversely, failure. This knowledge is essential to build the appropriate intervention strategies which are targeted, timely, and effective. For example, in fall 2007, the first Engineering Learning Community (ELC) will be initiated. This program will require all incoming freshman students to live together in a community environment. Identification of success factors can provide a significant tool to identify the most effective interventions, the students who will benefit most from these interventions, and accelerate the positive impact of the ELC on student performance.

As a starting point in improving retention, this research examined indicators of future engineering program success which may be evident in the high school record at the time of application for admission. The goal is to determine if relationships to later performance levels could be established. On the performance side, since the program's initiation in fall 2004, two years of student records were available. For the first entering class, currently juniors, GPA for the freshman and sophomore years were available. For the second entering class, currently sophomores, only the GPA from the freshman year was available.

On the predictor side, as a component of the admissions process, the department had maintained several benchmarks of high school performance for the students in the program. The most consistently documented, quantitative set of high school performance data includes weighted HS GPA, un-weighted or core HS GPA, position in high school class, SAT-Math, and SAT-Verbal scores. These prediction factors are discussed in detail in the methodology section of this paper.

This paper presents the first analysis targeted at addressing these research questions:
- Can GPA for freshman and sophomore engineering students be predicted?
- If so, what factors are statistically significant?
- Based on this information, what strategies can be developed to improve GPA performance and enhance retention results?

This paper addresses the first two questions above. Future publications will continue to explore these questions and provide results of the longitudinal study under way to examine the third question.

Methodology and Results

This section examines the results of predicting student GPA using quantifiable variables tracked in program applicant data. A multivariable, linear regression model is employed with five predictor variables described in equation (1), where GPA is the dependent variable, A_0 is a constant, the b_i terms are multipliers, and the x_i terms are the five explanatory or predictor variables. The predictor variables and the dependent variable are more completely described in Table 1.

$$GPA = A_0 + b_1x_1 + b_2x_2 + b_3 x_3 + b_4 x_4 + b_5 x_5 \qquad (1)$$

Using a step wise regression approach, starting with the full prediction model including all five explanatory variables, GPA at the end of the first and second year for the 2004 entering class and GPA for the end of the first year for the 2005 entering class were predicted using the linear model described in equation (1). Predictors were removed in sequential steps based on least significance as a model factor, as indicated by a t distribution score. Equation (1) was also tested with $A_o = 0$, that is with no constant in the prediction equation. Details of the results are described in the following sections.

Table 1: Summary of Regression Model Variables

Dependent Variable	Description	Scale
GPA- grade point average	Cumulative grade point average of freshman and sophomore engineering students at the end of spring semester.	Continuous from 0 to 4.
Predictor variables	Description	Scale
Core GPA,	Core high school grade point average as reported at time of admission	Continuous 0 to 4, normalized on 0 to 1 scale
Weighted GPA	High school grade point average as reported at time of admission, but adjusted based on difficulty of courses such as advanced placement.	Continuous 0 to 4, normalized on 0 to 1 scale. May exceed 4.0 and consequently 1.0.
Class rank	Position in the high school graduating class as specified at time of admission.	Continuous 0 to 1 scale with 1 being top of class and 0 being bottom of class
SAT- V	Verbal SAT score as reported at time of admission	Continuous in interval 200-800, normalized on 0 to 1 for model.
SAT-M	Math SAT score as reported at time of admission	Continuous in interval 200-800, normalized on 0 to 1 for model.
SAT Total	Total of verbal and math SAT scores as reported at time of admission.	Continuous in interval 400-1600, normalized on 0 to 1 for model.

Prediction of first year GPA for class entering 2004

Table 2 and Table 3 contain the statistical results for the stepwise regression analysis for this first data set. Table 2 describes the results with a constant (i.e. $A_o \neq 0$) and Table 3 describes results without a constant (i.e. $A_o = 0$). On this first regression analysis, the R-squared values indicate the model with a constant factor resulted in a data fit which is inferior to the model without a constant. Since these results are consistent in both of the following cases, the results of the stepwise regression for the model with a constant are only presented in this case and not included in the next two.

In general, a t score of approximately ± 2 indicates a factor which is significant at the 95% confidence level. Several points can be made based on Tables 2 and 3:

- As noted, the R-squared values in Table 3 indicate the model without a constant explains a much higher proportion of the variation in the data and is a superior model to Table 2.
- SAT-M is the least significant factor and is the first variable removed from both models.
- Weighted HS GPA is the next least significant factor and was removed from the model after the second step.
- SAT –V remained in the model until after the third step but was removed due to its lack of significance in both models.
- Core GPA, followed by position in class, are the strongest predictors of first year GPA for this group of 27 students in both models. In fact, core GPA in Table 3 is the only prediction factor with a statistical significance above 90% in the last two steps in both models.

Table 2: Class of 2004, First Year GPA Prediction Model with Constant

	Stepwise Regression Sequence				
	1	2	3	4	5
R-Sq (% variation explained)	22.84	22.13	17	12.25	6.2
Constant (Ao)	0.17383	-0.02543	0.65142	0.15163	1.12487
Core GPA	**5.1**	**4.9**	**6.5**	**5.1**	**1.8**
T-Value	**1.45**	**1.42**	**2.07**	**1.75(+90%)**	**1.29**
Wtd. GPA	3.8	3.8			
T-Value	1.2	1.2			
Class rank	**-4.6**	**-4.6**	**-3.1**	**-2.3**	
T-Value	**-2.04**	**-2.04**	**-1.64**	**-1.29 (~80%)**	
SAT-V	-2.1	-2.2	-1.8		
T-Value	-1.28	-1.39	-1.15		
SAT-M	-0.7				
T-Value	-0.44				

Note: Two most significant predictor variables in bold. Confidence value in (%) noted

Table 3: Class of 2004, First Year GPA Prediction without Constant

	Stepwise Regression Sequence				
	1	2	3	4	5
R-Sq (%)	93.39	93.33	93.83	92.42	
No constant					
Core GPA	**5.25**	**4.83**	**7.48**	**5.41**	**3.21**
T-Value	**1.61**	**1.57**	**3.19 (+99%)**	**4.02 (+99%)**	**16.92**
Wtd. GPA	3.9	3.7			
T-Value	1.35	1.32			
Class rank	**-4.8**	**-4.5**	**-3.4**	**-2.4**	
T-Value	**-2.38**	**-2.38**	**-1.98 (~95%)**	**-1.65 (~90%)**	
SAT-V	-2	-2.2	-1.6		
T-Value	-1.31	-1.46	-1.08		
SAT-M	-0.7				
T-Value	-0.44				

Note: Two most significant predictor variables in bold. Confidence value in (%) noted

Table 4 examines the results of stepwise regression using a combined SAT-Total score instead of individual Math (SAT-M) and Verbal (SAT-V) scores. The results of the analysis related to Table 2 and 3 did not change. SAT–Total was not a significant predictor of GPA and was removed from the model after the second step. Once again, core GPA and class rank were the most significant predictors. Since these results are consistent in the following models, SAT-Total will no longer be discussed as a potential model component to replace individual SAT scores.

Table 4: Class of 2004, First Year GPA Prediction without Constant and with SAT-Total

	Stepwise Regression Sequence			
	1	2	3	4
R-Sq (%)	93.29	92.78	92.48	
No constant				
Core GPA	**5.24**	**7.7**	**5.41**	**3.21**
T-Value	**1.63**	**2.88**	**4.02 (+99%)**	**16.92**
Wtd. GPA	3.8			
T-Value	1.32			
Class rank	**-4.5**	**-3.4**	**-2.4**	
T-Value	**-2.34**	**-1.92**	**-1.65 (~90%)**	
SAT - Total	-2.7	-1.8		
T-Value	-1.4	-0.99		

Note: Two most significant predictor variables in bold. Confidence value in (%) noted

Prediction of Second Year GPA for Class Entering 2004

Table 5 presents the results of the step wise regression model predicting the GPA at the end of the second year. Of the original 27 students with a complete set of factor data included in the first year GPA prediction model, only 16 finished both semesters of the second year with an equivalent set of course work completed.

- Similar to the first year GPA model, SAT-M is the least significant factor and is the first variable removed from the model.
- Core HS GPA was the second factor to be removed from the model. In the first year case, it was the most important factor.
- In step three, weighted GPA, class rank, and SAT-V were statistically significant predictors in that order.
- Removal of SAT-V in step four did not impact R-squared significantly and the simpler model including weighted GPA and class rank is justified.

Table 5: Class of 2004, Second Year GPA Prediction without Constant

	Stepwise Regression Sequence			
	1	2	3	4
R-Sq (%)	97.3	97.3	97.09	96.28
No constant				
Core GPA	2.9	3.2		
T-Value	0.7	0.96		
Wtd. GPA	**8.8**	**8.7**	**10**	**6.6**
T-Value	**3.3**	**3.45**	**4.7 (+99%)**	**5.36 (+99%)**
Class rank	**-7.2**	**-7.3**	**-6.2**	**-4.1**
T-Value	**-3.19**	**-3.5**	**-3.57 (+99%)**	**-2.82 (+99%)**
SAT-V	-3.7	-3.7	**-2.7**	
T-Value	-2	-2.1	**-1.9 (~95%)**	
SAT-M	0.2			
T-Value	0.13			

Note: Two most significant predictor variables in bold. Confidence value in (%) noted

<u>Prediction of First Year GPA for Class Entering 2005</u>

Table 6 contains the model analysis for the first year GPA of the second entering class. In comparison with Table 3 (containing first year results for the first class) we find:
- Similar to Table 3, core GPA and class rank are the most significant predictors. However, both of these predictors did not have as large a t distribution value in Table 6, step 4 as in Table 3, step 4.
- SAT-M and SAT-V were not significant predictors as in Table 3. However, these predictors changed positions in the step wise removal sequence between the two tables.
- Weighted GPA was not significant in either table and was removed after the second step in both cases.

Table 6: Class of 2005, First Year GPA Prediction without Constant

	Stepwise Regression Sequence				
	1	2	3	4	5
R-Sq (%)	92.49	92.39	92.33	92.17	
No constant					
Core GPA	**2.65**	**2.69**	**1.67**	**2.16**	**3.11**
T-Value	**1.31**	**1.36**	**1.5**	**2.51 (+95%)**	**16.68**
Wtd. GPA	-1.4	-1.6			
T-Value	-0.52	-0.63			
Class rank	**1.55**	**1.57**	**1.04**	**1.12**	
T-Value	**1.15**	**1.19**	**1.04**	**1.14 (~75%)**	
SAT-V	-0.5				
T-Value	-0.29				
SAT-M	1.3	1.2	0.7		
T-Value	0.94	0.92	0.69		

Note: Two most significant predictor variables in bold. P-value in (%) noted

Conclusions

This study investigated factors that predict success of freshman and sophomore engineering students, as measured by end of year GPA. A multivariable, linear regression model was employed with five predictor variables; core high school GPA, high school weighted GPA, high school class rank, SAT-Verbal and SAT-Math scores. These factors are commonly used predictors by universities to determine a student's acceptance status. However, the data that has been collected and analyzed from these two cohorts of students suggest that several of these factors do not predict academic success. Table 7 details the significance of each of these predictors on the GPA of students at the end of their first and second years as identified using a step wise regressions approach.

Review of Table 7 indicates that predictors which have been used for many years to screen applicants may not be as important as once thought.
- HS class rank was statistically significant in all three years of this study.
- Core HS GPA was a significant predictor for both entering classes for determination of GPA after their first year; yet this factor was insignificant in determining GPA at

the end of year two. Counter to this was weighted GPA which was insignificant for both first year cohorts; however, weighted GPA was significant in predicting GPA at the end of year two. One possible hypothesis is that students who pass standard high school courses with high grades can succeed in college in the first year; however, students who are challenged with advanced or AP courses in high school acquire a level of improved study skills which becomes important over two years of a difficult degree program such as engineering.

- SAT-Math was not a significant predictor of success and SAT verbal scores did not have a statistically significant impact on college success except in the case of the second year cohort. It appears from this study that SAT scores in general are not as important as many commonly believe. A possible hypothesis is that SAT scores are a snapshot and do not measure a student's determination or persistence to succeed.

Continued monitoring and analysis of future classes will provide more data that can either substantiate or contradict these initial results.

Table 7: Summary of Regression Results

	Class Entering 2004: First year GPA (Table 3)	Class Entering 2005: First year GPA (Table 6)	Class Entering 2005: Second year GPA (Table 5)
Core GPA	Most significant predictor, significant at +99%	Most significant predictor, significant at +99%	Removed from model after second step
Wtd. GPA	Removed from model after second step	Removed from model after second step	Most significant predictor at +99%
Class rank	Second most important predictor but significant at ~90%	Second most important predictor but only significant at ~75%	Second most important predictor but significant at + 99%
SAT-V	Removed after third step	First variable removed	Removed after third step but significant at 90% in step three
SAT-M	First variable removed	Removed after third step	First variable removed
SAT- total	Did not improve model performance when used to replace SAT-V and SAT-M		

One potential area for future research is investigation of the signs of the prediction factor coefficients in the model. An expectation in a case like this is that the signs would generally be positive, especially with the significant predictors. However, the results were mixed and Table 8 summarizes the results for the first step model (all factors included).

- Core HS GPA was positive in all cases and this appears to substantiate its importance as a predictor of success.
- In all three scenarios, the SAT-Verbal coefficient for predicting college GPA had a negative sign. This might suggest that students who have the skills related to higher scores on this test may be more inclined to withdraw from an engineering major and pursue other degree areas. Another interesting hypothesis involves the relationship of high SAT-Verbal scores and persistence to succeed in a difficult quantitative major.
- Class rank was significant in all three models but its sign was not consistent. Class rank is influenced by many factors such as the size of the high school and the quality of students within that school. A school with many college bound student might cause the class rank of an above average student to be lower than it would be in a small school with fewer college bound students.

Table 8: Signs of Predictor Variables in Full Model

	Class Entering 2004: First year GPA (Table 3)	Class entering 2005: first year GPA (Table 6)	Class entering 2005: second year GPA (Table 5)
Core GPA	Positive	Positive	Positive
Wtd. GPA	Positive	Negative	Positive
Class rank	Negative	Positive	Negative
SAT-V	Negative	Negative	Negative
SAT-M	Negative	Positive	Positive

Bibliography

1. Takahira, S., Goodings, D., and Byrnes, B., "Retention and Performance of Male and Female Engineering Students: An Examination of Academic and Environmental Variables," *Journal of Engineering Education*, Vol. 88, 1998, pp. 297-304.
2. Ohland, Matthew W., Sharon A.Frillaman, Guili Zhang, Catherine E. Brawner, and Thomas K. Miller. "The Effect of an Entrepreneurship Program on GPA and Retention." *Journal of Engineering Education*, Oct. 2004, pp. 293-301.
3. Lackey, L., Lackey, W., Grady, H., Davis, M., "Efficacy of Using a Single, Non-Technical Variable to Predict the Academic Success of Freshmen Engineering Students," *Journal of Engineering Education*, 2003, Vol. 92, pp. 41-48.
4. French, B., Immekus, J., and Oakes, W., "An Examination Of Indicators Of Engineering Students' Success And Persistence," *Journal of Engineering Education,* 2005, pp.419-422.
5. Zhang, G., Anderson, T.J. , Ohland, M.W., and Thorndyke, B.R., "Identifying Factors Influencing Engineering Student Graduation: A Longitudinal and Cross-Institutional Study," *Journal of Engineering Education,* Vol. 93, 2004, pp. 313–320.
6. Lam, P., Doverspike, D., Mawasha, R., "Predicting Success In A Minority Engineering Program," *Journal Of Engineering Education*, Vol. 88 , 1999, pp. 256-267.
7. Budny, D., Lebold, W., and Bjedov, G., "Assessment of the Impact of Freshman Engineering Courses," *Journal of Engineering Education*, Vol 88, 1998, pp. 405-411.