Predicting Student Performance Using Multiple Regression

Ifte Choudhury

Texas A&M University
College Station, TX

Abstract

The purpose of this study was to examine the effects of overall academic capability of students on their performance in the Environmental Control Systems courses offered by the Department of Construction Science, Texas A&M University. It is indicated by a number of studies that student performance is affected by the overall academic capability of a student, measured by Grade Point Average. The study population consisted of the students who attended the Environmental Control System courses in Fall, Spring, and Summer semesters from 1997 through 2001. Sample size of the study was 329 students. Relevant data was collected from the Student Information Management System database of the university. In order to obtain a better fit of the data, a quadratic term of the independent variable was used in the model. The data was analyzed using a multiple regression analysis procedure. The findings generated from the analysis of the data indicated that overall academic capability and the major of a student have statistically significant effects on student performance in the Environmental Control Systems courses. The model conceived could be used for predicting student performance in these courses in order to enable the instructors to formulate teaching strategies geared toward helping the students at risk.

Statement of the problem

The Department of Construction Science within the College of Architecture at Texas A&M University offers Environmental Control Systems courses at an undergraduate level. The courses deal with building sub-systems such as HVAC (heating, ventilating, and air-conditioning), plumbing, lighting, and electrical wiring. Students completing these courses should be able to design these sub-systems and integrate them with the total building systems. It appeared that the students with higher academic capabilities performed better in the courses.

Studies on education indicate that any particular course grade is positively correlated with overall grade point average. Findings by Seymour et al.[1] reveal that most significant factor in predicting success in a business microcomputer course is the overall grade point average of a student. Similar findings have been reported by Rose et al.[2] in a study of student performance in an introductory psychology course. Vaidyanathan et al.[3] have shown that students with better GPA performed on computer-based simulations and regular tests in marketing classes than those with lower GPA scores.

Courses for all disciplines are geared toward meeting the requirements of the students enrolled in those disciplines. Non-major students taking those courses do not often perform well compared to the students for whom the courses have been designed[4].

It is, therefore, hypothesized that student performances in Environmental Control Systems courses, offered by the Department of Construction Science at Texas A&M University, have a positive relationship with their overall academic ability and academic major.

Methodology

Study Population

The study population consists of the students who registered for and actually attended the Environmental Control Systems course offered by the Department of Construction Science, Texas A& M University, and taught by the author in the following semesters:

1. Summer I 1997-2001
2. Spring 1998- 2001
3. Fall 2000.

The sample size was 329. The entities under study are the students who attended these classes. The unit of analysis is the student.

Data Collection

Data related to the grade point average of the students were collected from the Student Information Management Systems database of Texas A&M University. The student performance in the courses was collected from the author's own record.

Variables and their Operationalization

Student Performance (GRADE). Student performance is the actual academic performance of the student in the class. It was measured by the percentage of total numerical grade obtained by the student in the course.

Grade Point Average (GPA): It is the overall grade point average of the student. It ranges from 0 to 4. It was used as a proxy for overall academic ability of a student.

Academic Major (*MAJOR*): It is the discipline in which a student is enrolled. It is a dummy variable that took on a value of 1 for a construction science major and 0 for all others majors.

Analysis and Interpretation

Results

A simple linear regression technique was used to analyze the data. The following model was used for the analysis:

$$GRADE = \beta_0 + \beta_1 GPA + \beta_2 MAJOR \quad (1)$$

where *GRADE* = student performance in terms of numerical grade, *GPA* = grade point average of a student, *MAJOR* = academic major of a student, β_0 = intercept, β_1 = the coefficient of *GPA*, and β_2 = the coefficient of MAJOR.

The results of the analysis are shown in Table 1.

Table 1

Regression Analysis for GRADE using GPA and MAJOR as an independent variable

Variable	Intercept	Coefficient of GPA	T	$p>\|T\|$	Critical Value of $\|T\|$
Intercept	46.839		26.19	0.0001	
GPA		12.27	20.86	0.0001	1.29
MAJOR		2.27	4.36	0.0001	
F-value of the Model = 222.1	*p*>Model *F*= 0.0001	Model R^2 = 0.58 Adjusted model R^2 = 0.57			

Even though the coefficient of determination of the model was found to be moderately high and the F-value was statistically significant at 0.0001 level, the residual plot of the model indicated a lack of fit of the sample data (Figure 1).

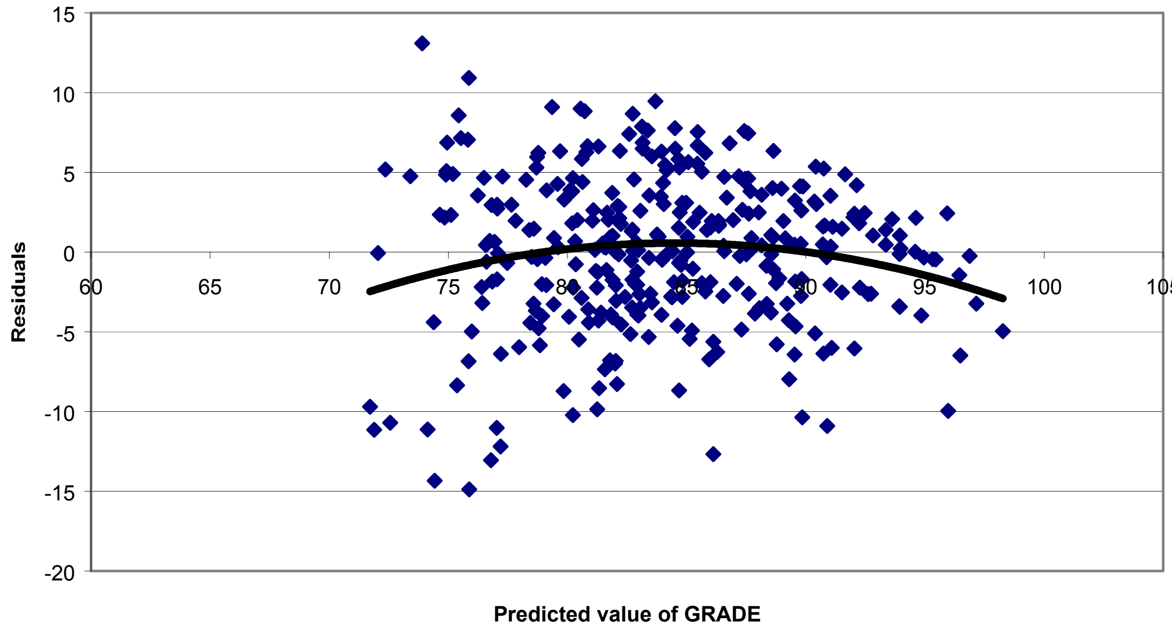

Figure 1. Residual plot 1

Another multiple regression analysis was performed using quadratic term of *GPA* along with the linear term. The following model was used for the purpose:

$$GRADE = \beta_0 + \beta_1 GPA + \beta_2 GPASQ + \beta_3 MAJOR \qquad (2)$$

where *GRADE* = student performance in terms of numerical grade, *GPA* = grade point average of a student, *GPASQ* = quadratic term of *GPA*, β_0 = intercept, β_1 = the coefficient of *GPA*, and β_2 = the coefficient of *GPASQ*, and β_3 = the coefficient of *MAJOR*.

The results of the analysis are shown in Table 2.

Table 2

Regression Analysis for GRADE using GPA, and GPASQ, and MAJOR as independent variables

Variable	Intercept	Regression Coefficients	*T*	*p*>|*T*|	Critical Value of |*T*|
Intercept	10.76		1.10	0.2710	
GPA		37.15	5.59	0.0001	1.29
GPASQ		-4.17	-3.76	0.0002	
MAJOR		2.08	3.93	0.0001	
F-value of the Model = 158.73	*p*>Model *F*= 0.0001	Model R^2 = 0.59 Adjusted model R^2 = 0.59			

The results indicate a small increase in the predictive efficacy of the model with an R^2 of 0.59 with an adjusted R^2 of 0.59. The residual plot on the model indicated a better fit of the sample data (Figure 2).

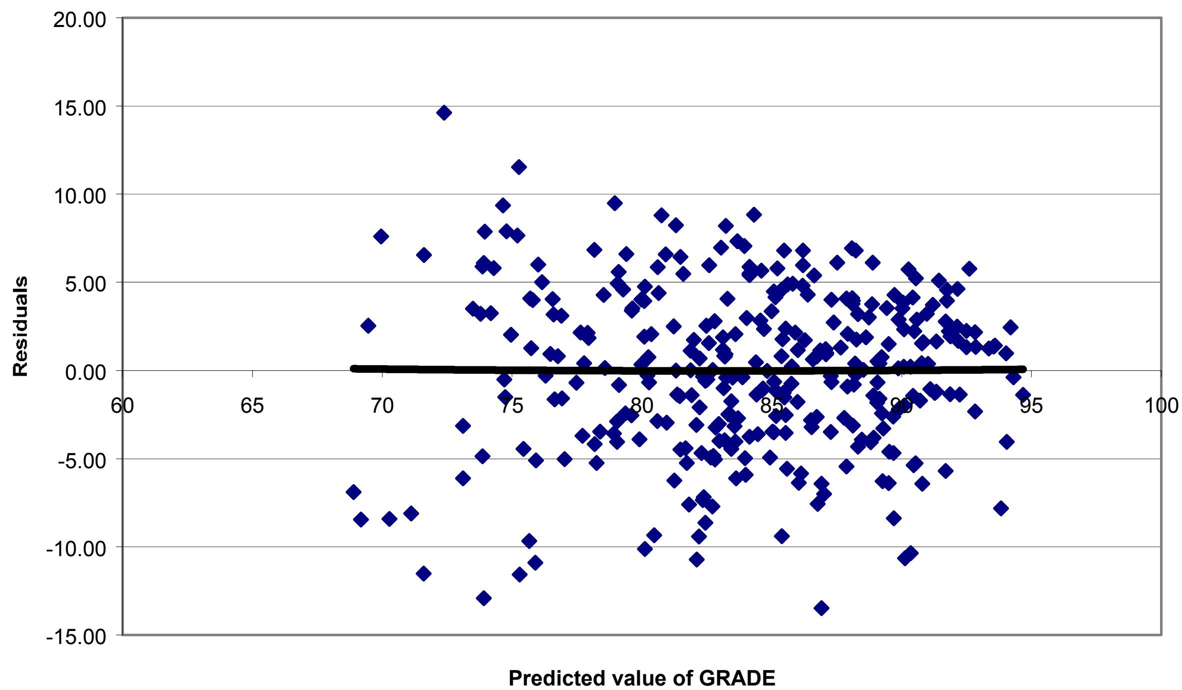

Figure 2. Residual plot 2

By plugging in the values of the intercept and regression coefficients, the prediction model for student performance in environmental control systems courses can now be rewritten as follows:

$$GRADE = 10.76 + 37.15*GPA - 4.17*GPASQ + 2.08*MAJOR \quad (3)$$

Interpretations

An important aspect of a statistical procedure that derives model from empirical data is to indicate how well the model predicts results. A widely used measure the predictive efficacy of a model is its coefficient of determination, or R^2 value. If there is a perfect relation between the dependent and independent variables, R^2 is 1. In case of no relationship between the dependent and independent variables, R^2 is 0. Predictive efficacy of this particular model was found to be moderately high with an R^2 of 0.59. The adjusted value of R^2 was also found to be the same.

The results indicated that student performance in Environmental Control Systems courses is positively related to the overall academic ability of a student, measured by GPA, and the academic major of a student, both at the level of significance of 0.0001. The quadratic term of GPA was also statistically significant at 0.0002 level. The F statistic of a model basically tests how well the model, as a whole, accounts for the dependent variable's behavior. The F-value of this particular model was found to be statistically significant at the 0.0001 level. The model predicts that a student with GPA of, say, 3 and with a major in construction science is likely to score 86.76 ($10.76 + 37.15*3 - 4.17*3^2 + 2.08*1$) percent in Environmental Control Systems courses. A non-major student with same GPA is likely to score 84.68 ($10.76 + 37.15*3 - 4.17*3^2 + 2.08*0$).

Discussions

The results of the statistical analysis indicated that students with higher overall capability tend to perform better in the Environmental Control Systems courses offered by the Department of Construction, Texas A&M University. Studies on education indicate that student performance in any particular course is positively correlated with overall grade point average. It was, therefore, only likely to find such a positive relationship with reference to the Environmental Control Systems.

Academic major of a student (*MAJOR*) was found to be a statistically significant predictor of student performance at the 0.0001 level. The course has been designed to meet the requirements for an undergraduate degree in Construction Science. Because of this reason it is possible that non-majors who take this course may lack an understanding of some of the concepts in order to appreciate the course and, therefore, are intimidated by the course content.

Conclusions

Based on the empirical data, a mathematical model has been developed to predict student performance in these courses. The model as a whole accounts quite well for the behavior of the dependent variable student performance. This is evident from the high *F*-value of the model that is statistically significant at the 0.0001 level. The predictive efficacy of the model is moderately high with an adjusted R^2 value of 0.59. It could be useful as a tool for identifying predicted poor performers in the environmental control systems courses. The process would enable the instructors to take remedial measures at an earlier stage to provide additional help to the groups at risk. The study will hopefully generate enough interest to do further research for deriving models for predicting student performance in other courses.

References

1. Seymour, J., Goings, D., & Vincent, A., 1994, “Factors contributing to success in a microcomputer course”, *Perceptual and Motor Skills*, Vol. 79, p. 1338.
2. Rose, R. J., Hall, C. W., & Bolen, L. M., 1996, “Locus control and college students' approaches to learning”, *Psychological Reports*, Vol. 79, pp. 163-171.
3. Vaidyanathan, R. & Rochford, L., 1998, “An exploratory investigation of computer simulations, student preference, and performance”, *Journal of Education for Business*, Vol. 73, pp. 144-149.
4. Choudhury, I., 2001, “Effects of Reciprocal Peer Tutoring on Student Performance in an Environmental Control Systems Course at an Undergraduate Level”, *Proceedings of the American Society for Engineering Education Annual Conference*, Albuquerque, New Mexico, June 24 – 27, 2001, on CD-ROM.

IFTE CHOUDHURY
Dr. Ifte Choudhury is an Assistant Professor in the Department of Construction Science. He has extensive experience as a consulting architect working on projects funded by the World Bank, Asian Development Bank, and some other Multilateral Development Banks. His areas of emphasis include housing, design of building sub-systems, alternative technology, and issues related to international construction.