Int. J. Engng Ed. Vol. 26, No. 4, pp. 1008–1017, 2010
Printed in Great Britain.

0949-149X/91 $3.00+0.00

Prediction of Student Academic Performance in an Engineering Dynamics Course: Development and Validation of Multivariate Regression Models*

SHAOBO HUANG and NING FANG
Department of Engineering and Technology Education, College of Engineering, Utah State University, Logan, UT 84322, USA. E-mail: ning.fang@usu.edu

The present study aims to develop a validated set of multivariate regression models to predict student academic performance in Engineering Dynamics—a high-enrollment, high-impact, and core engineering course. The models include eight predictor/independent variables that take into account student achievement before taking the course and student learning progression and achievement while taking the course. A total of 1,674 data points were collected from 186 undergraduate engineering students in two semesters. Four multivariate regression models were generated using different sample sizes of training datasets. The models were evaluated, validated, and compared using multiple criteria including R-square values, shrinkage, and prediction accuracy. The results show that the developed regression models have excellent predictability with 87–91% of the average prediction accuracy, and they have moderate predictability (46–66%) to generate good predictions (a good prediction is defined as a prediction that results in less than 10% of prediction error).

Keywords: multivariate regression; student academic performance; prediction accuracy; engineering dynamics

1. INTRODUCTION

ENGINEERING DYNAMICS is a high-enrollment, high-impact, and core engineering course that almost every mechanical, civil, aerospace, and biomedical engineering student is required to take [1–4].

1.1 Engineering dynamics and the importance of predicting student academic performance

Engineering dynamics is an essential basis and fundamental building block for advanced studies in many subsequent courses, such as vibration, structural mechanics, system dynamics and control, and machine and structural designs. However, many students fail this course because it covers a broad spectrum of foundational engineering concepts and principles, such as motion, force and acceleration, work and energy, impulse and momentum, and vibrations of a particle and of a rigid body [5–7].

Prediction of student academic performance has long been regarded as an important research topic in many academic disciplines because it benefits both teaching and learning [8–11]. Instructors can use the predicted results to identify the number of students who will perform well, average, or poorly in a class. For example, if the predicted results show that a particular group of students in a class would be "academically at risk," instructors could be proactive and take special measures to best accommodate the special needs of this group of students. Instructors may consider adopting active or inductive instructional strategies [12] or modifying the existing instructional strategies.

1.2 Multivariate regression

A variety of mathematical techniques have been employed in the development of a model to predict student academic performance. They include traditional statistical techniques such as multivariate regression [13–17] and modern data mining techniques such as neural networks [18], Bayesian networks [19], decision trees [20], and genetic algorithm [21]. Traditional statistical techniques, particularly the multivariate linear regression technique, have been most widely employed in educational research for at least two reasons. First, they do not require sophisticated mathematical skills for researchers to master, and therefore they are easy to understand and use [22, 23]. Second, traditional statistical techniques are most often associated with an explicit set of mathematical equations, allowing education researchers and practitioners to "see" how the predicted results are generated,

* Accepted 21 March 2010.

and thus the predicted results can be interpreted in a reasonable and meaningful way [24, 25].

One representative example is provided to show the application of the regression technique in engineering education research. Yousuf [26] developed a multivariate linear regression model to predict the academic performance of the students enrolled in Computer Science and Engineering Technology programs. A total of 125 students were surveyed using a three-part measuring instrument. The predictor/independent variables of Yousuf's model [26] included a student's career self-efficacy belief (that is a belief of one's ability to successfully perform a given task), math-SAT scores, high school GPA, and vocational interest. The results showed that self-efficacy contributed significant unique variance in prediction of the academic performance of students.

1.3 Objective, research questions, and scope of the present study

The objective of the present study is to develop a validated set of multivariate regression models to predict student academic performance in an Engineering Dynamics course. The research questions of the present study include:

What are the mathematical formulas of these regression models?

How accurate are these regression models when used to predict student academic performance?

The outcome/dependent variable (namely, the output Y) of the regression models is a student's score in the comprehensive final exam of the Dynamics course. The predictor/independent variables (namely, the inputs X_1, X_2, X_3, etc.) of the regression models include a student's:

X_1: Cumulative GPA
X_2: Grade earned in Engineering Statics (a prerequisite course)
X_3: Grade earned in Calculus I (a prerequisite course)
X_4: Grade earned in Calculus II (a prerequisite course)
X_5: Grade earned in Physics (a prerequisite course)
X_6: Score earned in Dynamics mid-exam #1
X_7: Score earned in Dynamics mid-exam #2
X_8: Score earned in Dynamics mid-exam #3

where X_1, X_2, X_3, X_4, and X_5 represent a student's prior achievement before the student takes the Dynamics course, and X_6, X_7, and X_8 are a direct representation of a student's learning progression and achievement in the Dynamics course during the semester before the student takes the comprehensive final exam of the course. The reason for selecting these particular variables as the predictor variables for the regression models will be explained in detail in a subsequent section of this paper: Research Method in the Present Study.

The scope of this paper is limited to investigation of the effects of cognitive factors on student academic performance and does not include a study of the effects of a student's non-cognitive factors (such as learning style, self-efficacy, motivation and interest, time devoted to learning, family background, race, and many others [27–29]), the instructor's teaching effectiveness and preparation [30], and teaching and learning environment [31].

1.4 Novelty and significance of present study

A variety of commonly-used literature databases were examined, including the Education Resources Information Center, Science Citation Index, Social Science Citation Index, Engineering Citation Index, Academic Search Premier, the ASEE annual conference proceedings (1995–2009), and the ASEE/IEEE Frontier in Education conference proceedings (1995–2009). The only paper on predictive modeling of student academic performance in the Engineering Dynamics course is the work done by Fang and Lu [20]. However, their work was primarily based on a decision tree (also called a classification tree) modeling approach, in which a collection of data records was split into branch-like segments using a sequence of "if-then" decision rules. These decision rules were then employed to predict the student's final grade (A, A-, B+, etc.) in the Dynamics course. For example, based on the developed decision tree model [20], if a student earned the following grades in their prerequisite courses: Statics = B AND Calculus II = C AND Physics = A, then the student would earn a grade of B in the Dynamics course.

A new set of research findings are generated from the present study based on extensive quantitative data (1,674 data points) collected from 186 undergraduate engineering students in two semesters. For example, four new regression models are generated by using different sample sizes of training datasets. It is revealed that the sample size of a training dataset does not significantly affect the average prediction accuracy of the developed models, but does affect the percentage of good predictions. The predicted results from the models developed in the present study can be used by instructors and students to improve teaching and learning in various ways.

1.5 Logic structure and contents of this paper

First, the research method of the present study is explained step by step. Second, data collection and pre-processing are described. Then, both descriptive and correlation analyses were performed to develop a fundamental insight into the collected first-hand data. Next, it is shown how multivariate regression models were developed using different sample sizes of training datasets, and how the models were evaluated, validated, and compared using multiple criteria including R-square values, shrinkage, and prediction accuracy. The limitations of the developed regression models are discussed. Conclusions are made at the end of the paper.

2. RESEARCH METHOD IN THE PRESENT STUDY

In the present study, the multivariate linear regression technique was employed to develop a set of models for predicting student academic performance in Engineering Dynamics. Data on student academic performance in two semesters (#1 and #2) were collected to develop and validate the models. The following paragraphs describe the research method step by step.

Step 1: Collect data on student academic performance in Semesters #1 and #2. Descriptive analysis and correlation analysis are performed to develop a fundamental understanding of the collected first-hand data.

Step 2: Randomly split the full dataset (128 students) collected in Semester #1 into a training dataset and a testing dataset. In this paper, the terms of "training" and "testing" are borrowed from the terms typically used in the neural network modeling technique. "Training" dataset is the samples employed to develop a regression model. "Testing" dataset is the samples employed to test the accuracy of the developed regression model. To investigate how a training dataset affects the prediction accuracy of its associated regression model, the following combinations of training datasets and testing datasets are employed:

- 30% (38 students) of the full dataset as the training dataset and the remaining 70% (90 students) as the testing dataset;
- 40% (51 students) of the full dataset as the training dataset and the remaining 60% (77 students) as the testing dataset;
- 50% (64 students) of the full dataset as the training dataset and the remaining 50% (64 students) as the testing dataset;
- 60% (77 students) of the full dataset as the training dataset and the remaining 40% (51 students) as the testing dataset.

Step 3: Develop a multivariate linear regression model based on each training dataset.

Step 4: Test each regression model developed in Step 3 using the corresponding testing dataset. Multiple criteria including R-square, shrinkage, and prediction accuracy are employed to test each model. Because both training and testing datasets are from the same Semester #1, Step 4 is also called the "*internal validation*" of the regression models.

Step 5: Apply the regression models developed in Semester #1 to the full dataset collected in Semester #2, and determine the prediction accuracy of each model. Because the models are applied to students in a *different semester* in the same university, Step 6 is also called the "*external validation*" of the regression models. In this paper, "external" does not mean a different university.

Eight variables (X_1, X_2, . . . , X_8) are selected as the predictor/independent variables of the regression models. The reasons for selecting these variables are given below:

X_1 (Cumulative GPA) is included because it is a comprehensive measurement of a student's cognitive level and problem-solving skills.

X_2 (Statics grade) is included because numerous concepts of Statics (such as free-body diagram, force equilibrium, and moment equilibrium) are employed in the Dynamics course.

X_3 and X_4 (Calculus I and II grades) are employed to measure a student's mathematical skills needed to solve calculus-based Dynamics problems.

X_5 (Physics grade) is used to measure a student's basic understanding of physical concepts and principles behind various Dynamics phenomena. Students at our university take either Physics 2200 (Elements of Mechanics) or Physics 2210 (General Physics).

X_6 (score of mid-term exam #1) measures student problem-solving skills on the first group of Dynamics topics: "kinematics of a particle" and "kinetics of a particle: force and acceleration."

X_7 (score of mid-term exam #2) measures student problem-solving skills on the second group of Dynamics topics: "kinetics of a particle: work and energy" and "kinetics of a particle: impulse and momentum."

X_8 (score of mid-term exam #3) measures student problem-solving skills on the third group of Dynamics topics: "planar kinematics of a rigid body" and "planar kinetics of a rigid body: force and acceleration."

The final exam (that is, the output Y of the regression models) is comprehensive and covers all the above-listed Dynamics topics as well as three additional topics that students learned after mid-term exam #3. The three additional topics include "planar kinetics of a rigid body: work and energy," "planar kinetics of a rigid body: impulse and momentum," and "vibration." The details of all these essential learning topics in Engineering Dynamics can be found in well-known Hibbeler's textbook [6].

3. DATA COLLECTION

Data on student academic performance were collected from a total of 186 students in two semesters: 128 students in Semesters #1 and 58 students in Semester #2. One of the authors was the instructor of the Dynamics course in both semesters. Table 1 shows student demographics.

As seen from Table 1, the majority of the 186 students were either from the mechanical and aerospace engineering major (50.5%) or from the civil and environmental engineering major (29.0%). The vast majority of students were male (85.5 %), and the female students accounted for 14.5%.

Table 1. Student demographics

	Major *			Sex	
	MAE	CEE	Other	Male	Female
Semester #1 (n = 128)	72 (56.3%)	34 (26.5%)	22 (17.2%)	108 (84.4%)	20 (15.6%)
Semester #2 (n = 58)	22 (37.9%)	20 (34.5%)	16 (27.6%)	51 (87.9%)	7 (12.1%)
Total (n = 186)	94 (50.5%)	54 (29.0%)	38 (20.5%)	159 (85.5%)	27 (14.5%)

* MAE: Mechanical and aerospace engineering. CEE: Civil and environmental engineering. Other: Biological engineering, general engineering, pre-engineering, undeclared majors, etc.

For each student, nine data were collected including the final exam score (Y) of the Dynamics course and the values of eight predictor/independent variables (from X_1 to X_8). For a two-semester total of 186 students, $186 \times 9 = 1{,}674$ data points were collected.

4. PRE-PROCESSING OF COLLECTED DATA

Collected data (Y, X_1, X_2, X_3, . . . , X_8) were initially in different scales of measurements: X_1 varied from 0 to 4; X_2, X_3, X_4, and X_5 varied from A to F (letter grades); and X_6, X_7, X_8, and Y varied from 0 to 100. Before using them to establish regression equations, the collected raw data must be pre-processed.

First, all letter grades in X_2, X_3, X_4, and X_5 were converted into the corresponding numerical values using Table 2, so linear regression models (other than logistic regression models) could be developed.

Second, the numerical values of all data were normalized, so each datum varied within the same range from 0 to 1, as shown in Table 3. The purpose of data normalization was to avoid the cases in which one variable received a high or low weight in its regression coefficient due to its initial low or large scale of measurements. The normalized value of data was calculated through dividing the initial value of the data by its maximum possible value in its same category. For instance, the maximum GPA that a student could receive is 4.00. Supposing one student earned a GPA of 3.55, the normalized GPA of that student would be $3.55 \div 4.00 = 0.8875$.

5. DESCRIPTIVE ANALYSIS

Table 4 shows the results of descriptive statistics of the normalized data for Semesters #1 and #2, respectively.

Table 2. Conversion of letter grades

Letter grade	A	A–	B+	B	B–	C+	C	C–	D+	D	F
Numerical value	4.00	3.67	3.33	3.00	2.67	2.33	2.00	1.67	1.33	1.00	0.00

Table 3. Normalization of collected raw data

Variables	Meaning	Initial value of data	Normalized value of data
X_1	Cumulative GPA	0–4 (numerical value)	0–1
X_2	Grade earned in Engineering Statics	Letter grade A, A–, B+, B, etc.	0–1
X_3	Grade earned in Calculus I	Letter grade A, A–, B+, B, etc.	0–1
X_4	Grade earned in Calculus II	Letter grade A, A–, B+, B, etc.	0–1
X_5	Grade earned in Physics	Letter grade A, A–, B+, B, etc.	0–1
X_6	Score earned in Dynamics mid-exam #1	0–100 (numerical value)	0–1
X_7	Score earned in Dynamics mid-exam #2	0–100 (numerical value)	0–1
X_8	Score earned in Dynamics mid-exam #3	0–100 (numerical value)	0–1
Y	Score earned in Dynamics final exam	0–100 (numerical value)	0–1

Table 4. Descriptive statistics of normalized data for Semester #1 and (Semester #2)

Variable	Minimum	Maximum	Mean	Standard deviation
Cumulative GPA	0.62 (0.51)	1.00 (0.99)	0.86 (0.81)	0.10 (0.11)
Engineering Statics	0.40 (0.33)	1.00 (1.00)	0.81 (0.67)	0.19 (0.21)
Calculus I	0.40 (0.42)	1.00 (1.00)	0.76 (0.76)	0.19 (0.19)
Calculus II	0.40 (0.42)	1.00 (1.00)	0.78 (0.73)	0.18 (0.20)
Physics	0.40 (0.19)	1.00 (1.00)	0.79 (0.74)	0.16 (0.19)
Mid-exam #1	0.27 (0.33)	1.00 (1.00)	0.79 (0.71)	0.16 (0.18)
Mid-exam #2	0.44 (0.38)	1.00 (1.00)	0.78 (0.78)	0.14 (0.14)
Mid-exam #3	0.47 (0.40)	1.00 (1.00)	0.85 (0.81)	0.12 (0.15)
Final exam	0.32 (0.33)	1.00 (1.00)	0.72 (0.69)	0.17 (0.16)

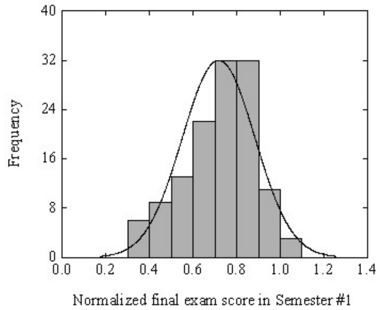


Fig. 1. Histogram of students' normalized final exam scores in Semester #1 (n = 128).

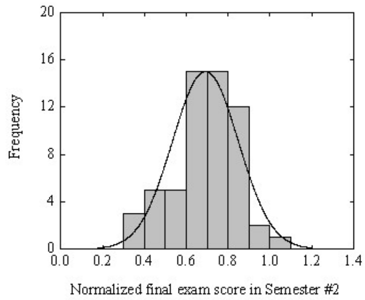


Fig. 2. Histogram of students' normalized final exam scores in Semester #2 (n = 58).

As seen from Table 4, students in Semester #2 had a lower mean and a higher standard deviation in most variables. For example, compared to students in Semester #1 as a whole, students in Semester #2 had a lower cumulative GPA, a lower Statics score, a lower mid-exam #3 score, and a higher standard deviation in GPA, Statics, and mid-exam #3 score.

There is a semester-to-semester variation in the student body. The above finding implies that students in Semester #2 (as a whole) did not perform as well as students in Semester #1, and that students in Semester #2 were more diverse in their academic performance. A comparison between student majors (refer to Table 1) in the two semesters shows that Semester #1 had more students majoring in mechanical engineering than did Semester #2. Thus, Semester #2 provided a different, "external" case to validate the applicability of the regression models developed from the data collected in Semester #1. Figures 1 and 2 further show the histograms of students' normalized final exam scores in the Dynamics course in the two semesters.

6. CORRELATION ANALYSIS

Correlation analysis aims to study how and to what extent two variables are related to each other. The correlation coefficient measures the degree of linear association between two variables. In the present study, correlation analysis was performed to investigate to what extent each predictor/independent variable (i.e., the inputs X_1, X_2, X_3, etc.) related to the outcome/dependent variable (i.e. the output Y or the final exam score of Dynamics). Correlation analysis was performed on all students in the two semesters, and the results are shown in Tables 5 and 6.

As seen from Tables 5 and 6, a statistically significant co-relationship ($p < 0.01$ or $p < 0.05$) exists between the final exam score of Dynamics and each of the eight predictor/independent variables for both semesters with only one exception, that is, the co-relationship between the final exam score of Dynamics and the Calculus I score. This latter co-relationship was not statistically significant for Semester #1 (correlation coefficient $r = 0.035$) but was statistically significant for Semester #2 ($r = 0.270$ at $p < 0.05$). In order to develop a general predictive model to cover as many cases as possible, it was decided to include the Calculus I score as a predictor/independent variable in the predictive models.

7. MULTIVARIATE REGRESSION MODELS FOR PREDICTING STUDENT ACADEMIC PERFORMANCE

Data collected in Semester #1 were employed to develop four regression models using four different sample sizes of training datasets: 30%, 40%, 50%,

Table 5. Correlation coefficients for Semester #1

	Cumulative GPA	Engineering Statics	Calculus I	Calculus II	Physics	Mid-exam #1	Mid-exam #2	Mid-exam #3	Final exam
Cumulative GPA	1.000	0.695**	0.194*	0.668**	0.416**	0.475**	0.468**	0.298**	0.448**
Engineering Statics		1.000	0.108	0.477**	0.360**	0.446**	0.418**	0.347**	0.346**
Calculus I			1.000	0.200*	0.190*	0.023	0.020	–0.123	0.035
Calculus II				1.000	0.375**	0.365**	0.266**	0.186*	0.267**
Physics					1.000	0.246**	0.234**	0.207*	0.335**
Mid-exam #1						1.000	0.437**	0.358**	0.461**
Mid-exam #2							1.000	0.421**	0.370**
Mid-exam #3								1.000	0.550**
Final exam									1.000

** Correlation is significant at the 0.01 level (2-tailed).
* Correlation is significant at the 0.05 level (2-tailed).

Table 6. Correlation coefficients for Semester #2

	Cumulative GPA	Engineering Statics	Calculus I	Calculus II	Physics	Mid-exam #1	Mid-exam #2	Mid-exam #3	Final exam
Cumulative GPA	1.000	0.628**	0.329*	0.619**	0.512**	0.569**	0.578**	0.598**	0.636**
Engineering Statics		1.000	0.284*	0.424**	0.326*	0.502**	0.555**	0.603**	0.730**
Calculus I			1.000	0.411**	0.371**	0.233	0.379**	0.288*	0.270*
Calculus II				1.000	0.448**	0.416**	0.396**	0.568**	0.408**
Physics					1.000	0.358**	0.297*	0.425**	0.377**
Mid-exam #1						1.000	0.421**	0.530**	0.582**
Mid-exam #2							1.000	0.430**	0.672**
Mid-exam #3								1.000	0.666**
Final exam									1.000

** Correlation is significant at the 0.01 level (2-tailed).
* Correlation is significant at the 0.05 level (2-tailed).

and 60% of the full dataset. First, the students' final exam scores (maximum: 100) were divided into different levels: 100-90, 89-80, 79-70, 69-60, and below 59. Then, the training dataset was randomly chosen from 30%, 40%, 50%, or 60% of the full dataset at each level to ensure the training dataset was a good representation of all students' performance in the class. The results are shown in columns 3–6 in Table 7.

The mathematical formula of each regression model is expressed as:

$$Y_1 = -0.429 + 0.567X_1 - 0.233X_2 - 0.040X_3 + 0.050X_4 + 0.281X_5 + 0.258X_6 + 0.122X_7 + 0.334X_8 \quad (1)$$

$$Y_2 = -0.380 + 0.520X_1 - 0.006X_2 + 0.213X_3 + 0.051X_4 + 0.079X_5 + 0.084X_6 - 0.055X_7 + 0.585X_8 \quad (2)$$

$$Y_3 = -0.309 + 0.556X_1 - 0.194X_2 + 0.002X_3 - 0.028X_4 + 0.102X_5 + 0.251X_6 - 0.070X_7 + 0.591X_8 \quad (3)$$

$$Y_4 = -0.334 + 0.500X_1 - 0.201X_2 - 0.021X_3 - 0.057X_4 + 0.154X_5 + 0.281X_6 + 0.053X_7 + 0.540X_8 \quad (4)$$

Each regression model was evaluated using the following four criteria that involved the use of either training or testing datasets:

(1) R-square value that represents the percentage that a model can explain its output based on a training dataset. The higher the R-square value, the better the model.
(2) Shrinkage value that indicates the loss of generalization ability (predictability) when a model is applied to other samples (i.e., testing datasets in this case). Shrinkage is calculated as

$$\text{Shrinkage} = R^2 - \left[1 - \frac{n-1}{n-k-1} \cdot \frac{n-2}{n-k-2} \cdot \frac{n+1}{n}(1 - R^2)\right] \quad (5)$$

where n is the number of students, and k is the number of predictor variables in the model. The lower the shrinkage value, the better the model. For example, a shrinkage value of 0.226 means that if a regression model is applied to a new set of samples, there would be a 22.6% of reduction in the generalization ability of the model.
(3) Average prediction accuracy for final exam scores, which indicates on average, how well a model predicts final exam scores of students in the Dynamics course. The average prediction accuracy for final exam scores is calculated as

Average prediction accuracy for final exam scores =

$$\frac{1}{n} \cdot \sum_{i=1}^{n} \left| \frac{P_i - A_i}{A_i} \right| \times 100\% \quad (6)$$

where n is the total number of predictions; P_i is the predicted final exam score of the i^{th} student

Table 7. Regression models based on the data in Semester #1

		Using a training dataset	Internal validation using a testing dataset		
Regression model No.	**Sample size (training dataset / full dataset)**	**R-square**	**Shrinkage**	**Average prediction accuracy (%)**	**Percentage (%) of good predictions among all predictions**
1	30%	0.415	0.385	89.2	51.7
2	40%	0.498	0.226	87.7	61.0
3	50%	0.403	0.200	90.7	65.6
4	60%	0.430	0.152	89.4	60.8

in the class (i = 1 to n); and A_i is the actual final exam score of the i^{th} student. The higher the average prediction accuracy, the better the model. For example, an average prediction accuracy of 88% (or 30% in a different case) means that if the model is applied to 50 students in a class, on average, the predicted final exam score for each student will be within 12% (or 70% in the different case) of the actual final exam score of the student.

(4) Percentage of good predictions among all predictions. This percentage is calculated as the number of good predictions divided by the total number of predictions. In the present study, a good prediction is defined as the prediction in which the predicted value is within 10% of the actual value. The higher the percentage of good predictions, the better the model.

As seen from Table 7, the average prediction accuracy varies within only 3% (min: 87.7% for Model No. 2; max: 90.7% for Model No.3) among the four regression models.

This means the sample size of the training dataset does not have a significant effect on the average prediction accuracy. However, the sample size does affect the percentage of good predictions, the latter percentage varying from 51.7% for Model No. 1 to 65.6% for Model No. 3. In terms of both the average prediction accuracy and the percentage of good predictions, Model No. 3 is apparently the best model among the four models. This is because the training dataset employed to develop Model No. 3 was probably the best representation of academic performance of all students in the class.

In addition, the following regression model was developed using only the first five variables X_1–X_5 (that is, excluding mid-term exam scores X_6, X_7 and X_8) and using 50% of the full dataset collected in Semester #1 as the training dataset.

$$Y_5 = 0.131 + 0.756X_1 - 0.100X_2 - 0.128X_3 - 0.011X_4 + 0.152X_5 \quad (7)$$

For the above equation, the average prediction accuracy is 88.7%, and the percentage of good prediction is 57.8%. For Model No. 3, the average prediction accuracy is 90.7%, and the percentage of good predictions is 65.6% (refer to Table 7). Therefore, without including variables X_6, X_7 and X_8, the percentage of good predictions would be reduced by approximately 8%.

Table 8. Validation of developed regression models

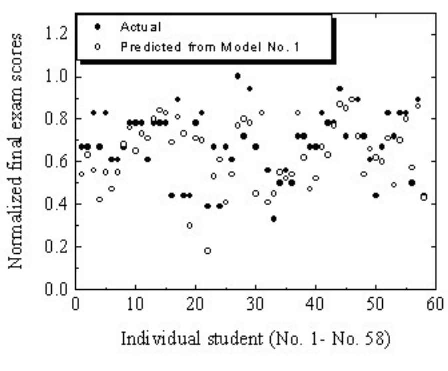
Model No.	Average prediction accuracy (%)		Percentage (%) of good predictions	
	Internal validation	External validation	Internal validation	External validation
1	89.2	88.1	51.7	46.6
2	87.7	87.3	61.0	48.3
3	90.7	89.8	65.6	56.9
4	89.4	90.1	60.8	56.9

8. EXTERNAL VALIDATION OF REGRESSION MODELS

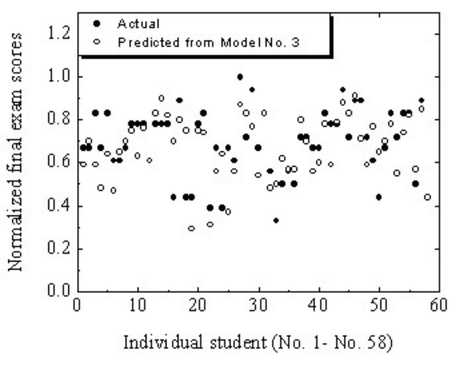
The totally different datasets collected in Semester #2 were employed to validate the regression models that were developed from the datasets collected in Semester #1. This is called external validation, and the results are shown in Table 8. To make comparisons clearer, some of the internal validation results provided in Table 7 is also included in Table 8.

As seen from Table 8, generally speaking, the prediction accuracy of the developed regressions models reduces when they were applied for external validation. The average prediction accuracy could be reduced up to 1.1% (for Model No. 1). However, the percentage of good predictions could be reduced up to 12.7% (for Model No. 2). Based on the results of both internal and external validation, it can be concluded that the developed regression models have excellent predictability with 87–91% of the average prediction accuracy, but they have moderate predictability (46–66%) to generate good predictions (again, a good prediction is defined as a prediction that results in less than 10% of prediction error).

As two representative examples, Figs 3a and 3b

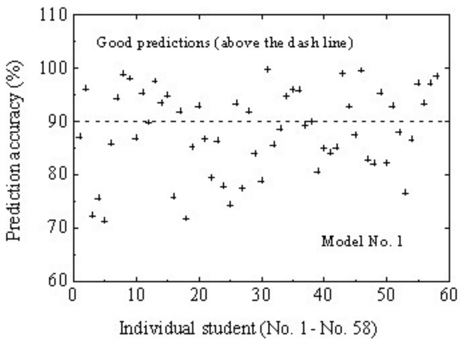


(a)

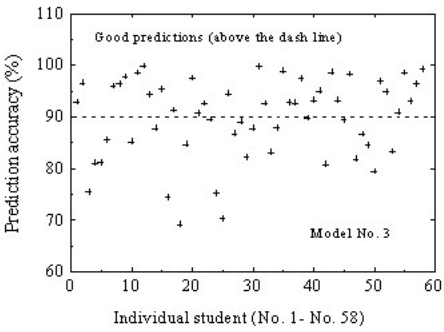


(b)

Fig. 3. Comparison of normalized final exam scores for 58 students in Semester #2 using a) Model No. 1 and b) Model No. 3.



(a)



(b)

Fig. 4. Comparison of prediction accuracy for 58 students in Semester #2 using a) Model No. 1 and b) Model No. 3.

show the predicted and actual normalized final exam scores for each of the 58 students in Semester #2, based on Models No. 1 and 3, respectively. In Fig. 3, each student was associated with two data points: a solid symbol for the actual final exam score and an open symbol (above or below the solid symbol in the same vertical line) for the predicted final exam score. The prediction accuracy for each student as well as good predictions are shown in Fig. 4.

9. DISCUSSION

Numerous factors affect student academic performance in teaching and learning processes. The analysis performed above shows that developed regression models can be employed to predict the average final exam score of all students in the Dynamics class with high prediction accuracy of around 90%. In other words, on average, the predicted final exam score for each student will be within approximately 10% of the actual final exam score. Nevertheless, the percentage of good predictions is only moderate from 46% to 66%. Two reasons are discussed in the following paragraphs.

First, the regression models developed in the present study are based only on cognitive factors, and do not take into account numerous non-cognitive factors such as a student's motivation and interest, learning style, self-efficacy [27–29]; the instructor's teaching effectiveness and preparation [30], and teaching and learning environment [31]. As Bransford, Brown, et al. [32] pointed out that human learning is a very complex process that involves numerous activities in three domains: cognitive (mental skills and knowledge), affective (feelings, emotion, and attitude), and psychomotor (manual and physical skills) domains. Xu et al. [33] also emphasized the importance of uncertain human factors in student learning processes such as classroom activities, instructor-student interactions, and student–student interactions in group activities. All these uncertainties can be clearly seen from the dispersion of the predicted results in Figure 3. The predicted results are therefore meaningful only in a statistical sense.

Second, the developed regression models do not take into account uncertain factors (such as a sudden health-related issue of a student, and personal or family emergency) that affect student learning and therefore student academic performance. A detailed examination of the values of all variables (X_1, X_2, X_3, . . . , X_8, and Y) for each student in two semesters shows that there were quite a few "outlier" students whose final exam scores could not be reasonably explained based on their prior performance and learning progression in the Dynamics course. For example, one student with cumulative GPA of 3.44 and increasingly good scores in mid-term exams in the Dynamics course (exam #1: 67/100; exam #2: 76/100; exam #3: 87/100) earned a final exam score of only 44/100. That student turned out to be very sick the day just before the final exam. This uncertainty factor was not included in the regression models.

In addition, it must be pointed out that the regression models in this paper were developed based on the data collected at our university. They can be employed as a general tool or guideline to predict student academic performance in the Dynamics course, so they can benefit both teaching and learning. When extending the regression technique to a particular institution of higher learning, it is suggested that data be collected data on student academic performance there to develop a corresponding regression model. This will ensure that the regression model best represents teaching and learning at that particular institution.

10. CONCLUSIONS

In the present study, four different sample sizes of training datasets have been employed to develop four linear regression models to predict student academic performance in an Engineering Dynamics course. The inputs (predictor/independent variables) of the models include a student's cumulative GPA; grades earned in Engineering Statics, Calculus I, Calculus II, and Physics; as well as scores earned in Dynamics mid-exams #1, #2, and #3. The output (outcome/dependent variable) of the models is a student's final exam score in the Dynamics course.

Descriptive analysis shows that students in Semester #2 (as a whole) did not perform as well as students in Semester #1, and that students in Semester #2 were more diverse in their academic performance. Thus, Semester #2 provided an excellent "external" case to validate the applicability of the regression models developed from the data collected in Semester #1.

Correlation analysis shows that a statistically significant co-relationship ($r = 0.27$ to 0.73, $p < 0.01$ or $p < 0.05$) exists between a student's final exam score in Dynamics and each of the eight predictor/independent variables for both semesters with only one exception.

Multiple criteria have been employed to evaluate the developed regression models, including R-square, shrinkage, the average prediction accuracy, and the percentage of good predictions. It is revealed that the sample size of training dataset does not have a significant effect on the average prediction accuracy. However, the sample size does affect the percentage of good predictions. Model No. 3, which was developed based on 50% of the full dataset, is the best among the four models in terms of both the average prediction accuracy and the percentage of good predictions. This is because the training dataset employed to develop Model No. 3 was probably the best representation of academic performance of all students in the class.

The results of both internal and external validation show that the developed regression models have excellent predictability with 87–91% of the average prediction accuracy, and they have moderate predictability (46–66%) to generate good predictions. Without including variables X_6, X_7 and X_8 (mid-term exam scores of students), the percentage of good predictions of the model would be reduced by approximately 8%. Non-cognitive factors, such as a student's motivation and interest, learning style, self-efficacy; the instructor's teaching effectiveness and preparation, and teaching and learning environment, will be included in future modeling efforts in order to improve the percentage of good predictions by the models.

REFERENCES

1. B. Self and R. Redfield, New approaches in teaching undergraduate dynamics. *Proceedings of the 2001 American Society for Engineering Education Annual Conference & Exposition*, Albuquerque, NM 2001.
2. M. B. Rubin and E. Altus, An alternative method for teaching dynamics. *International Journal of Engineering Education*, **16**, 2000, pp. 447–456.
3. R. Kumar and M. Plummer, Using contemporary tools to teach dynamics in engineering technology. *International Journal of Engineering Education*, **13**, 1997, pp. 407–411.
4. P. J. Cornwell, Teaching dynamics using modern tools. *Computers in Education*, **6**, 1996, pp. 18–24.
5. N. Fang, An electronic classroom response system for an engineering dynamics course: student satisfaction and learning outcomes. *International Journal of Engineering Education*, **25**, 2009, 1059–1067.
6. R. C. Hibbeler, *Engineering Mechanics Dynamics* (12th edition), Pearson Prentice Hall, Upper Saddle River, NJ 2009.
7. A. Bedford and W. Fowler, *Engineering Mechanics Dynamics* (5th edition), Prentice Hall, Upper Saddle River, NJ 2009.
8. R. A. Thompson and Byron L. Zamboanga, Academic aptitude and prior knowledge as predictors of student achievement in introduction to psychology. *Journal of Education Psychology*, **96**, 2004, pp. 778–784.
9. S. R. Ting, Predicting academic success of first-year engineering students from standardized test scores and psychosocial variables. *International Journal of Engineering Education*, **17**, 2001, pp. 75–80.
10. J. L. Holland and R. C. Nichols, Prediction of academic and extra-curricular achievement in college. *Journal of Educational Psychology*, **55**, 1964, pp. 55–65.
11. L. Cohen, Predicting academic success in an engineering college and suggestions for an objective evaluation of high-school marks. *Journal of Engineering Education*, **37**, 1946, pp. 381–384.
12. M. Prince, Does active learning work? A review of the research. *Journal of Engineering Education*, **93**, 2004, pp. 223–231.
13. J. H. McMillan and S. Schumacher, *Research in Education: Evidence-Based Inquiry* (7th Edition), Prentice Hall, Upper Saddle River, NJ 2009.
14. M. N. R. Ayan and M. T. C. Garcia, Prediction of university students' academic achievement by linear and logistic models. *The Spanish Journal of Psychology*, **11**, 2008, pp. 275–288.
15. C. M. Marsh, M. A. Vandehey and G. M. Diekhoff, A comparison of an introductory course to SAT/ACT scores in predicting student performance. *The Journal of General Education*, **57**, 2008, pp. 244–255.
16. T. C. Gilmer, An understanding of the improved grades, retention and graduation rates of STEM majors at the academic investment in math and science (AIMS) program of Bowling Green State University (BGSU). *Journal of STEM Education*, **8**, 2007, pp. 11–21.
17. E. H. Thomas and N. Galambos, What satisfies students? mining student-opinion data with regression and decision tree analysis. *Research in Higher Education*, **45**, 2004, pp. 251–269.

18. P. K. Imbrie, J. J. Lin, K. Reid and A. Malyscheff, Using hybrid data to model student success in engineering with artificial neural networks. *Proceedings of the Research in Engineering Education Symposium*, July 7-10, Davos, Switzerland 2008.
19. N. T. Nghe, P. Janecek and P. Haddawy, A comparative analysis of techniques for predicting academic performance. *Proceedings of the 37th ASEE/IEEE Frontiers in Education Conferences*, October 10-13, Milwaukee, WI 2007.
20. N. Fang and J. Lu, A decision tree approach to predictive modeling of student performance in engineering dynamics. *International Journal of Engineering Education*, **36**, 2010, pp. 87–95.
21. B. Minaei-Bidgoli, D. A. Kashy, G. Kortemeyer and W. F. Punch, Predicting student performance: an application of data mining methods with an educational web-based system. *Proceedings of the 33rd ASEE/IEEE Frontiers in Education Conferences*, November 5–8, Boulder, CO 2003.
22. J. W. Osborne, Prediction in multiple regression. *Practical Assessment, Research and Evaluation*, **7**, 2000, 13–21.
23. J. Cohen, P. Cohen, S. G. West and L. S. Aiken, *Applied Multiple Regression/Correlation Analysis for the Behavioral Sciences* (3rd edition), Lawrence Erlbaum Associates, Inc., Publishers, Mahwah, NJ 2003.
24. B. H. Cohen, *Explaining Psychological Statistics* (2nd edition), John Wiley & Sons, Inc., New York, NY 2000.
25. S. I. Green, Student assessment precision in mechanical engineering courses. *Journal of Engineering Education*, 94, 2005, pp. 273–278.
26. A. Yousuf, Self-efficacy and vocational interests in the prediction of academic performance of students in engineering technology. *Proceedings of the 2000 American Society for Engineering Education Annual Conference & Exposition*, June 18-21, St. Louis, MS 2000.
27. S. Ransdell, Predicting college success: the importance of ability and non-cognitive variables. *International Journal of Educational Research*, **35**, 2001, pp. 357–364.
28. T. J. Tracey and W. E. Sedlacek, Noncognitive variables in predicting academic success by race. *Measurement and Evaluation in Guidance*, **16**, 1984, pp. 171–178.
29. R. Riding and S. Rayner, *Cognitive Styles and Learning Strategies: Understanding Style Differences in Learning and Behavior*, David Fulton Publishers, London, UK 1998.
30. Y. Chen and L. B. Hoshower, Student evaluation of teaching effectiveness: an assessment of student perception and motivation. *Assessment & Evaluation in Higher Education*, **28**, 2003, pp. 71–88.
31. E. Graaff, G. N. Saunders-Smits and M. R. Nieweg, *Research and Practice of Active Learning in Engineering Education*, Palls Publications, Amsterdam, Holland 2005.
32. J. D. Bransford and A. L. Brown et al. *How People Know,* National Academy Press, Washington, DC 2000.
33. S. S. Xu, F. F. Ye and W. J. Li, Research on freshmen's autonomous learning competence. *Proceedings of the International Symposium on Total Engineering Education*, October 23–25, Shanghai, China 2009.

Shaobo Huang is a Ph.D. in Engineering Education student in the Department of Engineering and Technology Education in the College of Engineering at Utah State University (USU). With BS and M.S. degrees in electrical engineering, her area of interest focuses on the predictive modeling of student academic performance and problem solving skills in engineering courses. She is a recipient of the USU Presidential Fellowship.

Ning Fang is an Associate Professor in the College of Engineering at Utah State University, USA. He teaches Engineering Dynamics. His areas of interest include computer-assisted instructional technology, curricular reform in engineering education, the modeling and optimization of manufacturing processes, and lean product design. He earned his Ph.D., M.S., and BS degrees in mechanical engineering and is the author of more than 60 technical papers published in refereed international journals and conference proceedings. He is a Senior Member of the Society for Manufacturing Engineering and a member of the American Society of Mechanical Engineers. He is also a member of the American Society for Engineering Education and a member of the American Educational Research Association.