

# Modeling of student academic achievement in engineering education using cognitive and non-cognitive factors

Bothaina A. Al-Sheeb, A.M. Hamouda and Galal M. Abdella  
*College of Engineering, Qatar University, Doha, Qatar*

## Abstract

**Purpose** – The retention and success of engineering undergraduates are increasing concern for higher-education institutions. The study of success determinants are initial steps in any remedial initiative targeted to enhance student success and prevent any immature withdrawals. This study provides a comprehensive approach toward the prediction of student academic performance through the lens of the knowledge, attitudes and behavioral skills (KAB) model. The purpose of this paper is to aim to improve the modeling accuracy of students' performance by introducing two methodologies based on variable selection and dimensionality reduction.

**Design/methodology/approach** – The performance of the proposed methodologies was evaluated using a real data set of ten critical-to-success factors on both attitude and skill-related behaviors of 320 first-year students. The study used two models. In the first model, exploratory factor analysis is used. The second model uses regression model selection. Ridge regression is used as a second step in each model. The efficiency of each model is discussed in the Results section of this paper.

**Findings** – The two methods were powerful in providing small mean-squared errors and hence, in improving the prediction of student performance. The results show that the quality of both methods is sensitive to the size of the reduced model and to the magnitude of the penalization parameter.

**Research limitations/implications** – First, the survey could have been conducted in two parts; students needed more time than expected to complete it. Second, if the study is to be carried out for second-year students, grades of general engineering courses can be included in the model for better estimation of students' grade point averages. Third, the study only applies to first-year and second-year students because factors covered are those that are essential for students' survival through the first few years of study.

**Practical implications** – The study proposes that vulnerable students could be identified as early as possible in the academic year. These students could be encouraged to engage more in their learning process. Carrying out such measurement at the beginning of the college year can provide professional and college administration with valuable insight on students' perception of their own skills and attitudes toward engineering.

**Originality/value** – This study employs the KAB model as a comprehensive approach to the study of success predictors. The implementation of two new methodologies to improve the prediction accuracy of student success.

**Keywords** Retention, Prediction, Academic achievement, Engineering students

**Paper type** Research paper

## 1. Background and introduction

Universities are among the most important community institutions that are dedicated to the growth and development of their respective nations. The future of our nation is largely dependent on the technological and knowledge advancements offered by the field of engineering. Lin *et al.* (2008) and Wall (2010) emphasized the influence of the development of engineering and technology on the future of society. The document prepared by Wall (2010) recommends that a full measure of engineering capacity needs to be considered to improve



the quality of life in the developing world. Universities around the world face the pressure to meet local and global demands for highly qualified engineers.

Universities are tasked with satisfying the demand for qualified engineers. The National Academy of Engineering (2004) emphasized the role of engineering schools on the preparation and development of future engineers to satisfy the demands of the rapidly changing workforce. Many educational leaders have intensively focused on the challenge of recruiting and retaining students in the science, technology, engineering and mathematics (STEM) majors in general and in the engineering majors in particular. This was evident in the work by Jørgensen and Valderrama (2016). STEM-related job areas are expected to expand over the coming decades. Therefore, studying the success factors for the first- and second-year engineering students would be a valuable contribution to this field.

Students in engineering majors face a broad range of challenges during the first year of the course, and these challenges lead to attrition. Astin (1993) found that these challenges, specifically those faced by engineering students during their first year, led to lower college grade point average (GPAs). This concern was further supported by researchers such as Sax (2008) and Veenstra *et al.* (2009). These studies indicated that the demanding nature of engineering curricula causes many students to either shift to a different major, leave the university altogether or take a longer time to graduate. According to the report by the US National Science Foundation (2007), only 60 percent of the students who joined the engineering major graduated within six years. For this reason, it becomes important to focus on the critical-to-success factors for undergraduate engineering students in their first and early years of study at a four-year public university, i.e., Qatar University, Doha, Qatar.

### 1.1 *Conceptual/theoretical framework*

This study was based on a number of well-known theories in the education field. Tinto's (1993) interactionist theory forms an important part of the theoretical background for this study. The main idea of Tinto's model is "integration." He claimed that the persistence and retention of students can be predicted by the extent to which students are socially and academically integrated within the institution. We can relate this concept to some of our critical-to-success factors such as commitment to the college and a degree, scholastic self-regulation and time-management skills. Other factors related to student social integration include social connectedness, communication and teamwork skills.

The other theory that greatly affected this study is the expectancy-value model of achievement (Eccles and Wigfield, 2002). Some of the factors involved in this theory are motivation, career and scholastic self-efficacy and confidence in math and science skills. Using path analysis and structural equation modeling, Eccles and Wigfield found that "students' competence-related beliefs predict children's subsequent grades in mathematics and English beyond that predicted by their previous grades."

### 1.2 *Uniqueness of this study*

This study aims to approach and explore those critical-to-success factors through the lens of the knowledge, attitudes and behavioral skills (KAB) model which was first introduced to the education field by Schrader and Lawless (2004). In this study, the KAB model is implemented as a comprehensive approach to the study of the factors (both cognitive and non-cognitive) which are critical to the success of undergraduate engineering students. "KAB" refers to the combination of knowledge, attitudes and behavior. In this study, knowledge or the "cognitive factor" refers to the previous knowledge the student possess from high school education. The set of non-cognitive factors, under attitudes and behavioral skills, were mostly inspired by *Engineering Success* written by Schiavone (2001). Schiavone (2001) referred to these non-cognitive factors as the "study and survival skills" which most engineering students need in order to strive and thrive in the engineering major. Based on

both of these studies, this study determined the non-cognitive factors to be examined in this study and designed an instrument – “Building Blocks of Excellence” – to provide a quantified measure of student self-ratings for the two elements from the KAB model; “attitudes” and “behavioral skills.” This study is the first to explore the KAB model in the context of predicting students’ academic performance.

This study also aims to examine the efficiency of combining a classical variable-selection technique with ridge regression, a more recent technique. The study used two models. In the first model, exploratory factor analysis (i.e. principal component analysis) is used to reduce the dimension of the factors to be included in the ridge regression. The second model uses regression best-subset method to select the variables to be used for ridge regression. The combination of these techniques has not yet been explored in the education field. The efficiency of each of these models in terms of the  $R^2$  value and mean-squared error (MSE) is discussed in the Results section of this study.

## 2. Literature review

### 2.1 Cognitive and non-cognitive factors in the prediction of students’ academic outcomes

The persistence and academic success of college students are two different, yet inseparable, aspects affected by various factors and by combinations of factors (Lotkowski *et al.*, 2004). Researchers have confirmed that academic factors cannot be used alone to predict student performance (Astin and Astin, 1992). They have proposed the prospect of using non-cognitive factors as an alternative to standardized tests such as the scholastic aptitude test (SAT). A combination of high school grade point average (HSGPA) and standardized tests accounts for approximately 25 percent of the variance when predicting the first year’s GPA (Robbins *et al.*, 2004). We believe that more variance in the SAP could be explained through other non-cognitive factors.

To become a well-rounded student, students’ attitudes and other personality factors are as important as the knowledge dimension of the education process. Besterfield-Sacre *et al.* (1997) reported that education is an “aggregate of both cognitive (content knowledge and technical skills) and affective (attitudes) processes.” The non-cognitive factors are gaining momentum in the recent higher-education initiatives. According to a report of the Higher Education Academy, UK (Thomas *et al.*, 2017), “It is the human side of higher education that comes first – finding friends, feeling confident, and above all, feeling a part of your course of study and in the institution – that is the necessary point for academic success” (p. 1). This report confirms that non-cognitive factors, e.g., student’s sense of belonging, social connectedness and self-efficacy are the starting points for success and retention initiatives.

On a different perspective, information gathered about student perceptions of their attitudes and skill-related behaviors would be valuable to higher-education professionals. This information may support higher-education institutions in guiding resource allocation and program development for first-year engineering students (French *et al.*, 2003). Besterfield-Sacre *et al.* (1999) reported that students’ evaluation of their attitudes could provide educators with an important insight that would facilitate their understanding of student success and persistence.

Through a thorough examination of the literature related to predicting student academic performance in STEM majors, three streams of research can be found. Several studies relied on cognitive factors such as; high school exam scores (De Winter and Dodou, 2011), HSGPA and admission test scores (Jiang and Freeman, 2011; Andrews, 2014) and scores on standardized tests such as the ACT and the SAT (Kauffmann *et al.*, 2007; French *et al.*, 2005). Though high school grades and other admission tests might be good predictors for first-year GPA, Ackerman *et al.* (2013) argued that a great deal of individual difference variance remains unaccounted for.

The other stream of research focused only on non-cognitive factors; for example, Walton (2010) found that sequential construct of learning style and positive self, and focus constructs of

resilience were found to be predictors of student's persistence. The study by Stankov and Lee (2014) found that measures of motivation and self-efficacy had the highest predictive validity.

The third stream of studies considered both cognitive and non-cognitive factors in their attempts to understand the variance in student academic success and retention. An important study is the meta-analysis by Richardson *et al.* (2012), which found that performance self-efficacy had the largest correlation with GPA followed by high school GPA, ACT and grade goal. Gray *et al.* (2016) examined factors of learning as early indicators of students at-risk. Their study found that factors such as age, prior academic achievement, study efforts, self-efficacy and deep learning approach significantly predicted failing students.

The predictive validity of the combination of both cognitive and non-cognitive factors was generally agreed in the general education literature. Several studies from the engineering education field followed the same approach. Some of the early studies include those by Besterfield-Sacre *et al.* (1998), French *et al.* (2005), Burtner (2004, 2005), Bernold *et al.* (2007), Lin *et al.* (2008) and Veenstra *et al.* (2009). Non-cognitive factors that have been examined included variables such as emotional, social, psychosocial and study skill factors.

Recent contributions in the engineering field included the study by Jin *et al.* (2011). Their study found that while high school grades and SAT scores were better predictors of college GPA, and affective factors such as motivation and leadership were better predictors of first-year retention. In addition, Whitaker (2014) found that academic determination significantly contributed to students' GPAs above and beyond background academic preparation. The study by Hall *et al.* (2015) explored the impact of aptitudes and personality traits as well as HSGPA on student retention. Their study used the NEO Five-Factor Inventory and measure of locus of control to assess the personality trait. Hall *et al.* (2015) found that aptitudes tests such as high school GPA, SAT math and Calculus readiness placement test were significant predictors of student retention. In their study, "conscientiousness" from the big-five factors was the only significant personality factor.

From the literature reviewed, it is evident that different studies approached prediction of students' success differently and examined different aspects. There is a lack of a comprehensive framework, consistency or agreed upon structure that can be flexibly used by institutions aiming to study the cognitive and non-cognitive predictors of students' performance. To fill this gap in the literature, this study aims to study the success predictors through the comprehensive lens of the KAB model. To date, the KAB model has not been explored in this context in the field of predictive analytics.

## 2.2 Modeling techniques used for the prediction of students' performance

Several statistical modeling techniques have been used to estimate student success. The majority of such studies have used the classical linear regression or other types of regression. These approaches include logistic regression (Lin and Imbrie, 2014), multiple linear regression (Ting and Man, 2001; Virtanen *et al.*, 2013), hierarchical logistic regression (French *et al.*, 2005; Andrews, 2014), hierarchical multiple regression (Whitaker, 2014) and multinomial logistic regression (Hall *et al.*, 2015). Other statistical methods include path analysis (French *et al.*, 2003) and correlational analysis (Bernadin and McKendrick, 2015). Other classification methods that are employed in this area include discriminant analysis (Burtner, 2005) and artificial neural networks (Lin *et al.*, 2008; Huang, 2011). Classification methods can be used to predict categorical classes (discrete or nominal) but may not be used to model continuous-valued outcomes. The application of classical regression analysis is commonly used for this purpose.

However, multiple regression analysis is not usually the best option. The complexity of interpretation and over fitting are two common issues that often arise in high-dimensional data analysis. However, under this situation, the penalization-based regression techniques are more preferable over the classical regression techniques. The penalization-based regression techniques, such as the ridge and LASSO regression, use the sum of absolute

values of regression coefficients for penalizing the complexity of the regression model and then enhancing the accuracy of the prediction.

According to Chatterjee and Hadi (2015), when multicollinearity exists in a set of predictor variables, the ordinary least squares (OLS) of the regression coefficients tend to be unstable and can lead to inaccurate inferences. Ridge regression, a penalized form of the generalized linear model (GzLM), is an alternative to standard multiple regression when the predictive variables are highly correlated (multicollinearity). Allowing for a small bias to exist in the estimated coefficients, results in more meaningful estimates to be obtained. Ridge regression minimizes the error by introducing a penalty to the regression coefficients of the GzLM. This penalty causes the regression coefficients to shrink toward 0 value (but not to an exact 0). Hoerl and Kennard (1970) have demonstrated that some of these undesirable effects of multicollinearity can be reduced by using “ridge” estimates in place of the least squares estimates. The value of the shrinkage parameter,  $\lambda$ , is very critical to the quality of the ridge regression model. The value of this parameter can be specified so as to minimize the mean square error of the coefficient estimates of the model. Several methods have been developed over the last two decades for estimating the optimal value of  $\lambda$ . The cross-validation and the ridge trace plot are the most well-known among these methods. Moreover, the GzLM is also a method for relaxing the normality assumption in classical regression (Green and Silverman, 1993). Practitioners prefer GzLMs over the OLS and multiple linear analysis.

Ridge regression was used by Whittaker *et al.* (2000) to estimate the correlation of markers with a trait of interest. They have compared its performance to other techniques commonly used by researchers. Their study found that in all cases ridge performed better and produced smaller standard errors of selection response. They have concluded that ridge is a very stable procedure in the sense that small changes in the data do not produce large changes in the estimated regression coefficients. Erickson (1981) effectively used ridge regression to estimate directly lagged effects in marketing. Huang *et al.* (2008) proposed a method which combined ridge regression and hierarchical cluster analysis for investigating the effects of climate variations on bacillary dysentery incidence in northeast China. The effectiveness of ridge regression was also confirmed by El-Dereny and Rashwan (2011). They found that when multicollinearity exists all methods of ridge regression perform better than the OLS method.

### 2.3 Indicators and measures

The survey conducted in this study consisted of one dependent variable (student academic performance) and 11 independent variables: one cognitive variable (high school grade or HSG) and ten non-cognitive variables (five factors under attitudes and five factors under skill-related behaviors). In this study, the student academic performance is represented by the end-of-course GPA. The end-of-course GPA was used as a measure of student academic performance by many researchers in this field.

**2.3.1 Cognitive factors.** 2.3.1.1 HSG. A measure of student academic achievement or academic ability from high school.

From the literature reviewed, high school grades proved to be one of the most important cognitive factors that significantly predicted college GPA (Geiser and Veronica Santelices, 2007; Hiss and Franks, 2014).

**2.3.2 Non-cognitive Factors.** 2.3.2.1 Attitudes

- Motivation: motivation refers to a student’s desire and willingness to study engineering. This section is adapted from the Situational Motivation Scale, which represents a self-report measure of situational intrinsic motivation, identified regulation, external regulation and amotivation. This scale was developed by Deci and Ryan (1985, 1991) and was tested by Guay *et al.* (2000).

- Academic self-regulation: academic self-regulation is referred to as self-discipline. It was defined by Allen *et al.* (2008) as the extent to which a student appreciates work and tasks related to college and approaches them conscientiously. Studies by Pintrich and De Groot (1990) and Vrugt and Oort (2008) found that self-regulation is highly correlated with student GPA.
- Commitment to college and degree: a student's commitment to college and degree refers to the students' perception of the extent of their dedication and commitment to completing college and pursuing their degree.
- Scholastic and career self-efficacy: according to the definition provided by Bandura (1994), self-efficacy is defined as one's beliefs about his/her own capabilities to carry out their academic tasks efficiently. Several researchers including Robbins *et al.* (2004), Richardson *et al.* (2012) and Saltonstall (2013) found that student's self-efficacy is among the best predictors of student academic performance.
- Conscientiousness: in Le *et al.* (2005), conscientiousness was defined as "the extent to which a student is self-disciplined, achievement-oriented, responsible, and careful." The components of this section of the survey were adapted from the conscientiousness scale in the Big Five Personality Traits Test. According to several researchers such as Trapmann *et al.* (2007) and Poropat (2009), among the Big Five Personality Traits, conscientiousness is the only trait that has a significant correlation with academic achievement.

#### 2.3.2.2 Skill-related behavior

- Social connectedness: social connectedness is also referred to as social involvement. It refers to the extent to which a student participates in social activities or has social connections.
- Communication: Gore (2006) defined this term as "attentive to others' feelings and flexibility in resolving conflicts with others." Le *et al.* (2005) defined communication as "the ability to exchange information with others." This section is adapted from Duran's (1983) Communicative Adaptability Scale and Rubin and Martin's (1994) Interpersonal Communication Competence Scale.
- Confidence in math and science skills: this term was defined by Sheppard *et al.* (2010) as a student's perception of his/her proficiency in science, critical thinking, real-world problem-solving and computations. The statements in this section are adapted from the "mathematics self-efficacy and anxiety questionnaire" developed by May (2009).
- Teamwork-oriented skills: this variable measures the extent to which the student prefers to work within a team. A few elements of this section were adapted from the students' attitudinal success instrument proposed by Reid (2009).
- Time-management skills: this factor refers to the students' perceived ability to use and manage their time efficiently. Statements in this section were adapted from the "time-management questionnaire" developed by Britton and Tesser (1991).

#### 2.4 Study questions

The following questions guided this study:

- (1) How can a comprehensive framework be built to serve as a conceptual model for the prediction of students' academic performance?

- (2) To what extent can a group of cognitive and non-cognitive factors explain variability in students' final GPA?
- (3) Can using a selection or reduction technique optimize the prediction capacity of ridge regression?
- (4) Which reduction or selection technique is more efficient in the optimization of the prediction capacity of ridge regression?

### 3. Methodology

An instrument, called the "Building Blocks of Excellence," was developed and administered. The first section includes questions about the student's name, gender, employment status and number of credit hours registered. The second section is titled "attitudes," which represents student perceptions of five factors that are related to their attitudes toward engineering. The third part includes five factors about students' self-report or perceptions toward their own "skill-related behavior." The gained knowledge, which is the remaining component of the KAB approach, is represented by the student's high school final grade. Except for the demographics sections, students select a response based on a Likert scale (strongly agree = 5; agree = 4; neutral = 3; disagree = 2; strongly disagree = 1) for sections two through four. The factors corresponding to each section are shown in Table I.

The study instrument was administered to students in the College of Engineering in Qatar University. The first page of the survey included a letter to the students and a consent form. By signing the consent form, students granted permission to the researcher to obtain their GPA and other information from the Registration Department.

The survey targeted four courses of general requirements: CENG 106 Computer Programming, CENG 107 Engineering Skills and Ethics, GENG 300 Numerical Methods and GENG 200 Probability and Statistics for Engineers. Engineering undergraduates highly populate these courses. A total of 322 students answered the survey.

The study examined the correlation between course GPA and students' cognitive and non-cognitive factors. Preliminary analyses were performed using RStudio and SPSS packages to ensure that the measure met the required assumptions for the proposed statistical tests. Factors were measured for indicators of normality and outliers. Descriptive statistics were then generated to examine means, standard deviations and the minimum and maximum value for each predictive variable. Two methods of variable reduction and selection were performed. Each method was used before the ridge regression was carried out. The first method uses exploratory factor analysis, whereas the second method uses regression best-subset selection techniques.

| Section No. | Attitudes                           | No. of statements | Section No. | Skill-related behavior                | No. of statements |
|-------------|-------------------------------------|-------------------|-------------|---------------------------------------|-------------------|
| 1           | Motivation                          | 8                 | 6           | Communication                         | 9                 |
| 2           | Scholastic self-regulation          | 8                 | 7           | Social engagement                     | 9                 |
| 3           | Commitment to college and degree    | 7                 | 8           | Confidence in math and science skills | 7                 |
| 4           | Scholastic and career self-efficacy | 7                 | 9           | Teamwork-oriented skills              | 7                 |
| 5           | Conscientiousness                   | 7                 | 10          | Time-management skills                | 7                 |

**Table I.**  
Factors related to "attitudes" and "skill-related behavior"

#### 4. Data modeling and analysis

##### 4.1 Descriptive statistics

General descriptive statistics display an overview of the general demographics of participating students (Table II). The collected sample included a total of 144 (45.1 percent) first-year students and 175 (54.8 percent) second-year students. Male and female students accounted for 51.4 and 48.58 percent of the sample population, respectively.

Descriptive statistics were carried out for the 11 predictive variables. Table III displays the mean, standard deviation and minimum and maximum values of each of the predictive variables.

##### 4.2 Correlation matrix among variables and dependent variables

Correlations of our critical-to-success factors and student performance were established using the bivariate Pearson correlation, which determines the strength of linear relationships between two variables. Table IV summarizes the  $R^2$  for the critical-to-success factors with student performance or final GPA.

##### 4.3 Implementation of the proposed methodologies

4.3.1 *Method I: factor-analysis-based ridge regression analysis.* This study applied exploratory FA, as a dimension reduction technique, to reduce the variables used for the ridge regression model and explore the underlying structure of a large set of variables. The groupings or the components generated using the FA are used along with another two categorical variables (gender and group) in the ridge regression model.

We tested the assumptions for exploratory FA. According to Kim and Mueller (1978), the first assumption of FA is that the sample size is adequate. This assumption is measured on the basis of the Kaiser–Meyer–Olkin measure of sampling adequacy and the

|               | First-year students | Student group<br>Sophomore students | Total        |
|---------------|---------------------|-------------------------------------|--------------|
| <i>Gender</i> |                     |                                     |              |
| Male          | 77 (53.47%)         | 87 (49.7%)                          | 164 (51.4%)  |
| Female        | 67 (46.52%)         | 88 (50%)                            | 155 (48.58%) |
| Total         | 144 (45.1%)         | 175 (54.8%)                         |              |

**Table II.**  
General demographics of participants

| Factors                                    | Mean   | SD      | Min. | Max. |
|--|--------|---------|------|------|
| X1: High school GPA                        | 4.6090 | 0.20769 | 3.71 | 5.00 |
| X2: Motivation average                     | 3.9975 | 0.57866 | 1.81 | 5.00 |
| X3: Scholastic self-regulation             | 3.4348 | 0.54845 | 2.00 | 5.00 |
| X4: Commitment to college and degree       | 4.1211 | 0.63309 | 2.43 | 5.00 |
| X5: Scholastic and career self-efficacy    | 3.9024 | 0.52791 | 2.29 | 5.00 |
| X6: Conscientiousness average              | 3.6506 | 0.55857 | 2.00 | 5.00 |
| X7: Social engagement average              | 3.1829 | 0.59165 | 1.13 | 4.88 |
| X8: Communication skills average           | 3.4764 | 0.50838 | 2.44 | 4.89 |
| X9: Confidence in math and science average | 3.6594 | 0.64730 | 1.44 | 5.00 |
| X10: Teamwork average                      | 3.4174 | 0.49789 | 2.29 | 5.00 |
| X11: Time-management skills average        | 3.3683 | 0.52891 | 2.14 | 5.00 |

**Table III.**  
Descriptive statistics of student's performance factors

**Table IV.**  
Bivariate Pearson  
correlation matrix for  
student final GPA  
and 11 variables

|     | (Y)     | X1      | X2      | X3      | X4      | X5      | X6      | X7      | X8      | X9      | X10     | X11    |
|-----|---------|---------|---------|---------|---------|---------|---------|---------|---------|---------|---------|--------|
| Y   | 1       |         |         |         |         |         |         |         |         |         |         |        |
| X1  | 0.384** |         |         |         |         |         |         |         |         |         |         |        |
| X2  | 0.075   | -0.013  |         |         |         |         |         |         |         |         |         |        |
| X3  | 0.165** | 0.065   | 0.178** |         |         |         |         |         |         |         |         |        |
| X4  | 0.236** | 0.104   | 0.668** | 0.381** |         |         |         |         |         |         |         |        |
| X5  | 0.255** | 0.087   | 0.374** | 0.259** | 0.499** |         |         |         |         |         |         |        |
| X6  | 0.140*  | 0.121*  | 0.263** | 0.572** | 0.444** | 0.330** |         |         |         |         |         |        |
| X7  | 0.221** | 0.064   | 0.211** | 0.179** | 0.253** | 0.300** | 0.276** |         |         |         |         |        |
| X8  | 0.077   | -0.005  | 0.323** | 0.307** | 0.396** | 0.294** | 0.428** | 0.410** |         |         |         |        |
| X9  | 0.327** | 0.174** | 0.348** | 0.211** | 0.477** | 0.489** | 0.327** | 0.386** | 0.383** |         |         |        |
| X10 | -0.005  | 0.011   | 0.155** | 0.202** | 0.238** | 0.259** | 0.323** | 0.363** | 0.426** | 0.208** | 0.471** |        |
| X11 | 0.025   | -0.001  | 0.155** | 0.226** | 0.290** | 0.351** | 0.227** | 0.199** | 0.174** | 0.221** | 0.232** | 0.148* |

**Notes:** \*\*Significant at the 0.05 and 0.01 levels (two-tailed), respectively

*p*-value was (0.840). The second assumption is that correlations among factors are reliable. This assumption was checked using Bartlett's test of sphericity and the value of this test was (0.0). The third assumption is multicollinearity or singularity; we tested this assumption by checking the values of the tolerance and the variance inflation factor (VIF). Tolerance values were 0.79, 0.78, 0.78 and 0.96, and were all greater than 0.1. Values of the VIF (tolerance) were 1.259, 1.268, 1.181 and 1.04. Normality is not an assumption of FA. Using the eigenvalue criterion and interpretability, we extracted four components. As shown in Table V, the first four components had eigenvalues greater than one and they explained 65.546 percent of the variance in the data.

On the basis of the pattern matrix shown in Table VI, four groupings were established. We elected to suppress variables with a loading value of less than 0.4. According to the study of Guadagnoli and Velicer (1988), a factor with ten loadings greater than 0.4 is considered more stable for a sample size that is greater than 150. A title was assigned to each of the four components, as shown in Table VII.

In addition to two categorical variables (gender and group), the four components or the variables generated by the PCA were used in the ridge regression model, as shown in Table VIII.

**Table V.**  
Total variance  
explained

| Factors | Total | Initial eigenvalues |              | Extraction sums of squared loadings |               |              | Rotation sums of squared loadings |
|---------|-------|---------------------|--------------|-------------------------------------|---------------|--------------|-----------------------------------|
|         |       | % of variance       | Cumulative % | Total                               | % of variance | Cumulative % | Total                             |
| 1       | 3.917 | 35.608              | 35.608       | 3.917                               | 35.608        | 35.608       | 3.101                             |
| 2       | 1.181 | 10.736              | 46.344       | 1.181                               | 10.736        | 46.344       | 2.534                             |
| 3       | 1.068 | 9.705               | 56.049       | 1.068                               | 9.705         | 56.049       | 2.245                             |
| 4       | 1.045 | 9.496               | 65.546       | 1.045                               | 9.496         | 65.546       | 1.148                             |
| 5       | 0.886 | 8.052               | 73.597       |                                     |               |              |                                   |
| 6       | 0.668 | 6.076               | 79.674       |                                     |               |              |                                   |
| 7       | 0.600 | 5.456               | 85.130       |                                     |               |              |                                   |
| 8       | 0.530 | 4.822               | 89.952       |                                     |               |              |                                   |
| 9       | 0.430 | 3.910               | 93.862       |                                     |               |              |                                   |
| 10      | 0.400 | 3.633               | 97.495       |                                     |               |              |                                   |
| 11      | 0.276 | 2.505               | 100.000      |                                     |               |              |                                   |

**Note:** Extraction method: principal component analysis

To specify the optimal value of the tuning parameter  $\lambda$ , we used the ridge trace plotted by Hoerl and Kennard (1970). This method is a kind of trade-off between the bias and the variance of the coefficient estimates. This method suggests to visualize the changes in the GLM coefficients over a wide range of  $\lambda$  values and to select the value of  $\lambda$  when the changes in these coefficients are insignificant. The optimal range of  $\lambda$  is found to be between 0.1 and 0.55. A set of ten values of  $\lambda$  have been selected from this range, including extremes. The extreme values of the optimal range were intentionally included in order to maintain a small bias (see Table VIII). In Table VIII, the best ridge parameter value is 0.1.

|  | Component |       |       |       |
|--|-----------|-------|-------|-------|
|  | 1         | 2     | 3     | 4     |
| X1: High school grade                      |           |       |       | 0.954 |
| X2: Motivation average                     | 0.884     |       |       |       |
| X3: Scholastic self-regulation             |           |       | 0.900 |       |
| X4: Commitment to college and degree       | 0.797     |       |       |       |
| X5: Scholastic and career self-efficacy    | 0.660     |       |       |       |
| X6: Conscientiousness average              |           |       | 0.742 |       |
| X7: Social engagement average              |           | 0.771 |       |       |
| X8: Communication skills average           |           | 0.648 |       |       |
| X9: Confidence in math and science average | 0.568     |       |       |       |
| X10: Teamwork average                      |           | 0.798 |       |       |
| X11: Time management skills average        |           |       |       |       |

**Notes:** Extraction method: principal component analysis; rotation method: Oblimin with Kaiser normalization, rotation converged in ten iterations

**Table VI.**  
Pattern matrix

| Academic mindset                            | Social skills                     | Self-discipline                 | Prior knowledge              |
|---|-----------------------------------|---------------------------------|------------------------------|
| Motivation average (X2)                     | Social engagement (X7)            | Scholastic self-regulation (X3) | High school final grade (X1) |
| Commitment to college and degree (X4)       | Communication skills average (X8) | Conscientiousness average (X6)  |                              |
| Scholastic and career self-efficacy (X5)    | Teamwork average (X10)            |                                 |                              |
| Confidence in math and science average (X9) |                                   |                                 |                              |

**Table VII.**  
Scaled groupings based on factor analysis

| $\lambda$                     | 0.1   | 0.15  | 0.2   | 0.25  | 0.3   | 0.35  | 0.4   | 0.45  | 0.5   | 0.55  |
|-------------------------------|-------|-------|-------|-------|-------|-------|-------|-------|-------|-------|
| <i>Coefficient</i>            |       |       |       |       |       |       |       |       |       |       |
| $\beta_0$ : Intercept         | 2.96  | 2.96  | 2.95  | 2.95  | 2.94  | 2.94  | 2.94  | 2.94  | 2.93  | 2.93  |
| $\beta_1$ : Gender            | 0.08  | 0.08  | 0.08  | 0.08  | 0.08  | 0.08  | 0.08  | 0.07  | 0.07  | 0.07  |
| $\beta_2$ : Student group     | -0.26 | -0.25 | -0.24 | -0.23 | -0.22 | -0.22 | -0.21 | -0.20 | -0.20 | -0.19 |
| $\beta_3$ : Academic mindsets | 0.12  | 0.12  | 0.11  | 0.11  | 0.11  | 0.10  | 0.10  | 0.10  | 0.09  | 0.09  |
| $\beta_4$ : Social skills     | 0.01  | 0.01  | 0.01  | 0.01  | 0.01  | 0.01  | 0.02  | 0.02  | 0.02  | 0.02  |
| $\beta_5$ : Self-discipline   | -0.03 | -0.03 | -0.03 | -0.02 | -0.02 | -0.02 | -0.02 | -0.01 | -0.01 | -0.01 |
| $\beta_6$ : High school grade | 0.31  | 0.30  | 0.29  | 0.27  | 0.26  | 0.25  | 0.25  | 0.24  | 0.23  | 0.22  |
| MSE                           | 0.26  | 0.26  | 0.26  | 0.26  | 0.27  | 0.27  | 0.27  | 0.27  | 0.27  | 0.28  |

**Table VIII.**  
Ridge regression coefficients at different values of ridge parameter from 0.1 to 0.55

The MSE exhibited the smallest value (0.26) for this parameter. In this case, the final regression model is as follows:

$$\begin{aligned} \text{Student's GPA} = & 2.95905 + 0.0812297 \times \text{Gender} - 0.255536 \times \text{Student group} \\ & + 0.12479 \times \text{Academic mindset} + 0.00934197 \times \text{Social skills} \\ & - 0.0315418 \times \text{Self-discipline} + 0.310555 \times \text{High school grade.} \end{aligned} \quad (1)$$

The  $R^2$  statistics shown in Table IX indicate that the as-fitted model explains 34.59 percent of the variability in the students' final GPA. The adjusted  $R^2$  statistic is 31.922 percent. The standard error of the estimate shows the standard deviation of the residuals to be 0.51051. The mean absolute error (MAE) of 0.40449 is the average value of the residuals.

*4.3.2 Method II: best-subset-based ridge regression analysis.* The regression model selection procedure in the Statgraphics software was used to select the independent variables for use in the ridge regression model to predict our single quantitative dependent variable  $Y$  (students' final GPA). This method considers all possible regressions involving various combinations of the independent variables. This procedure compares regression models based on the MSE,  $R^2$ , adjusted  $R^2$  and the Mallows' Cp statistic. Here, we used the regression model selection procedure provided by Statgraphics to find the best subset of independent variables to use in the ridge regression model. Table X provides some examples of the regression models provided by the procedure. We selected the first model because it provided the best adjusted  $R^2$  value and the smallest MSE and Mallows' Cp values.

The selected variables in the first model were used in the ridge regression model. Table XI displays a summary of the regression coefficients corresponding to each value of

**Table IX.**  
Model results for  
ridge regression and  
residual analysis

| $\lambda$ | $R^2$ | $R^2$ (adjusted for df) | SE of Est. | Mean absolute error | Residual analysis |      |       |
|-----------|-------|-------------------------|------------|---------------------|-------------------|------|-------|
|           |       |                         |            |                     | MSE               | MAE  | MAPE  |
| 0.1       | 34.59 | 31.92                   | 0.51       | 0.40                | 0.26              | 0.40 | 15.90 |
| 0.15      | 33.24 | 30.52                   | 0.51       | 0.41                | 0.26              | 0.41 | 15.98 |
| 0.2       | 32.00 | 29.23                   | 0.51       | 0.41                | 0.26              | 0.41 | 16.05 |
| 0.25      | 30.86 | 28.04                   | 0.51       | 0.41                | 0.26              | 0.41 | 16.12 |
| 0.3       | 29.80 | 26.93                   | 0.52       | 0.41                | 0.27              | 0.41 | 16.20 |
| 0.35      | 28.81 | 25.90                   | 0.52       | 0.41                | 0.27              | 0.41 | 16.28 |
| 0.4       | 27.89 | 24.95                   | 0.52       | 0.41                | 0.27              | 0.41 | 16.36 |
| 0.45      | 27.03 | 24.05                   | 0.52       | 0.42                | 0.27              | 0.42 | 16.44 |
| 0.5       | 26.22 | 23.21                   | 0.52       | 0.42                | 0.27              | 0.42 | 16.54 |
| 0.55      | 25.46 | 22.42                   | 0.53       | 0.42                | 0.28              | 0.42 | 16.64 |

**Table X.**  
Some examples of the  
models generated by  
the best subset  
regression model

| MSE  | $R^2$ | Adjusted $R^2$ | Cp    | Included Variables |
|------|-------|----------------|-------|--------------------|
| 0.27 | 39.01 | 35.22          | 8.68  | ABCFIJKLM          |
| 0.27 | 39.40 | 35.19          | 9.76  | BCEFHIJKLM         |
| 0.27 | 38.55 | 35.18          | 7.76  | BCFIJKLM           |
| 0.27 | 39.38 | 35.17          | 9.81  | ABCFHJKLM          |
| 0.27 | 38.92 | 35.13          | 8.90  | BCFHIJKLM          |
| 0.27 | 39.74 | 35.10          | 10.97 | ABCFHIJKLM         |

**Notes:** A = Gender, B = Student group, C = X1, D = X2, E = X3, F = X4, G = X5, H = X6, I = X7, J = X8, K = X9, L = X10, M = X11, refer to Table III for more information on each variable

ridge parameter lambda ( $\lambda$ ). The best ridge parameters ranged from 0.1 to 0.55. The ridge parameter value of 0.1 clearly provided the best value of squared mean error as well as the highest  $R^2$  value.

A summary of the model results is shown in Table XII. In this case, the fitted regression model is as follows:

$$\begin{aligned} \text{GPA Final} = & -2.78441 + 0.0936042 \times \text{Gender} - 0.214495 \times \text{Student group} \\ & + 0.055593 \times \text{HSG} + 0.18744 \times \text{Commitment to college and degree} \\ & + 0.190047 \times \text{Social skills} - 0.160221 \times \text{Communication skills} \\ & + 0.232532 \times \text{Confidence in math and science} - 0.113575 \\ & \times \text{Teamwork skills} - 0.210142 \times \text{Time-management skills.} \end{aligned} \quad (2)$$

The  $R^2$  statistic indicates that the as-fitted model explains 35.5453 percent of the variability in the students' final GPA. The adjusted  $R^2$  is 31.5447 percent. The standard error of the estimate indicates the standard deviation of the residuals to be 0.523226. The MAE of 0.402426 is the average value of the residuals.

Finally, we use the results provided in Table XIII to draw a tentative conclusion about the efficiency of each model. Table XIII provides a comparative summary of the two models. The first model used PCA to reduce the dimensions of the variables into four components.

| $\lambda$                                    | 0.1   | 0.15  | 0.25  | 0.3   | 0.4   | 0.45  | 0.5   | 0.55  |
|--|-------|-------|-------|-------|-------|-------|-------|-------|
| <i>Coefficient</i>                           |       |       |       |       |       |       |       |       |
| $\beta_0$ : Intercept                        | -2.78 | -2.61 | -2.29 | -2.14 | -1.87 | -1.75 | -1.63 | -1.52 |
| $\beta_1$ : Gender                           | 0.09  | 0.09  | 0.09  | 0.09  | 0.09  | 0.08  | 0.08  | 0.08  |
| $\beta_2$ : Student group                    | -0.21 | -0.21 | -0.20 | -0.19 | -0.18 | -0.17 | -0.17 | -0.17 |
| $\beta_3$ : High school grade                | 0.06  | 0.05  | 0.05  | 0.05  | 0.04  | 0.04  | 0.04  | 0.04  |
| $\beta_4$ : Commitment to college and degree | 0.19  | 0.18  | 0.16  | 0.16  | 0.15  | 0.14  | 0.14  | 0.13  |
| $\beta_5$ : Social engagement                | 0.19  | 0.18  | 0.16  | 0.16  | 0.14  | 0.14  | 0.13  | 0.13  |
| $\beta_6$ : Communication skills             | -0.16 | -0.14 | -0.12 | -0.11 | -0.09 | -0.09 | -0.08 | -0.07 |
| $\beta_7$ : Confidence in maths and science  | 0.23  | 0.22  | 0.20  | 0.19  | 0.18  | 0.17  | 0.17  | 0.16  |
| $\beta_8$ : Teamwork skills                  | -0.11 | -0.11 | -0.10 | -0.09 | -0.09 | -0.08 | -0.08 | -0.08 |
| $\beta_9$ : Time-management skills           | -0.21 | -0.19 | -0.16 | -0.15 | -0.13 | -0.12 | -0.11 | -0.11 |
| MSE  | 0.27  | 0.28  | 0.28  | 0.28  | 0.29  | 0.29  | 0.29  | 0.29  |

**Table XI.**  
A summary of the regression coefficients corresponding to each value of ridge parameter Lambda ( $\lambda$ )

| $\lambda$ | $R^2$ | $R^2$ (adjusted for df) | SE of Est. | Mean absolute error | Residual analysis |      |       |
|-----------|-------|-------------------------|------------|---------------------|-------------------|------|-------|
|           |       |                         |            |                     | MSE               | MAE  | MAPE  |
| 0.1       | 35.55 | 31.54                   | 0.52       | 0.40                | 0.27              | 0.40 | 16.32 |
| 0.15      | 34.08 | 29.99                   | 0.52       | 0.40                | 0.28              | 0.40 | 16.44 |
| 0.2       | 32.75 | 28.58                   | 0.53       | 0.41                | 0.28              | 0.41 | 16.56 |
| 0.25      | 31.54 | 27.29                   | 0.53       | 0.41                | 0.28              | 0.41 | 16.67 |
| 0.3       | 30.43 | 26.11                   | 0.53       | 0.41                | 0.28              | 0.41 | 16.78 |
| 0.35      | 29.41 | 25.02                   | 0.53       | 0.41                | 0.28              | 0.41 | 16.90 |
| 0.4       | 28.46 | 24.02                   | 0.54       | 0.42                | 0.29              | 0.42 | 17.01 |
| 0.45      | 27.58 | 23.09                   | 0.54       | 0.42                | 0.29              | 0.42 | 17.12 |
| 0.5       | 26.76 | 22.22                   | 0.54       | 0.42                | 0.29              | 0.42 | 17.23 |
| 0.55      | 26.00 | 21.41                   | 0.54       | 0.42                | 0.29              | 0.42 | 17.33 |

**Table XII.**  
Model results for ridge regression at each value of ridge parameter

Labels were assigned to each of the four groupings and were saved as regression variables. The study used the four components generated by the PCA in the first ridge regression model. The second model used the regression best-subset procedure to reduce the number of variables to be used with the ridge regression. From the results in Table XIII, we conclude that although the second model provided a higher value of  $R^2$ , the PCA model provided a higher adjusted  $R^2$ , as well as lower MSE and mean absolute percentage error (MAPE) values. Though the differences in the values between the two models do not appear to be very large, some implications are made in the next section on the appropriate use for each model.

## 5. Discussion

### 5.1 Summary of this research

This study examined the potential impact of a combination of critical-to-success factors that were essentially based on the KAB approach. On the basis of the PCA, we generated four components: prior knowledge, academic mindset, social skills and self-discipline.

Academic mindsets were described by Farruggia *et al.* (2018) as directly and strongly affecting students' academic success. Motivation, as a component of mindset, is associated with college students' success in the studies carried out by French *et al.* (2003) and Kauffmann *et al.* (2007). However, Jin *et al.* (2011) suggested that academic motivation is a better predictor of engineering students' retention. Commitment to college and to completing a degree is also a good predictor of engineering success, in contrast to the study by Veenstra (2008). Confidence in math and science is a strong predictor of student performance, as confirmed by the studies of Lotkowski *et al.* (2004) and Robbins *et al.* (2004). Self-efficacy, the fourth element under academic mindset, is considered one of the best predictors of GPA. This relationship was evident in several studies including the meta-analysis by Robbins *et al.* (2004), Richardson *et al.* (2012) and Saltonstall (2013). The social skills construct had a moderate effect on students' GPAs, consistent with the study of Lotkowski *et al.* (2004). The last component, academic self-discipline, was a good predictor. It involves two variables: self-regulation and conscientiousness. Both of these variables have been demonstrated in the literature to strongly correlate with a student's success. The positive relationship between self-regulation and academic achievement was demonstrated by Pintrich and De Groot (1990) and Vrugt and Oort (2008). However, this notion is contrary to what was revealed by the study of Virtanen *et al.* (2013). Conscientiousness was also found in many studies to significantly affect student success (Trapmann *et al.*, 2007; Poropat, 2009).

We found that a combination of cognitive and non-cognitive factors contributes to approximately 35–40 percent of the variance in students' GPAs. This variance is greater than that found by many other studies in both general education and the engineering education field. While the non-cognitive factors accounted for approximately 25 percent of the variance in final GPA, cognitive variables contributed only 15 percent of this variance. Thus, we can conclude that non-cognitive factors are more important in predicting student academic performance.

The study used two prediction methods. Each method involved a combination of a classical data reduction/selection technique and the more recent technique of ridge regression. The first method used PCA to reduce the dimension of variables, whereas the

**Table XIII.**  
A comparative  
summary of the  
results obtained by  
the two models

|              | Selection method       | $R^2$ | Adjusted $R^2$ | Residual analysis |      |       |
|--------------|------------------------|-------|----------------|-------------------|------|-------|
|              |                        |       |                | MSE               | MAE  | MAPE  |
| First model  | FA (PCA)               | 34.59 | 31.92          | 0.26              | 0.40 | 15.90 |
| Second model | Regression best subset | 35.55 | 31.54          | 0.27              | 0.40 | 16.32 |

second used a regression best-subset model. The first model yielded a better adjusted  $R^2$  value and a lower MSE. The second model yielded a better  $R^2$  value.

We can conclude that through its KAB model, this study successfully explained the considerable variance in student success for engineering undergraduates. Moreover, adding a selection or dimension reduction technique further improved the prediction capacity of ridge regression. In addition, the predictive analytics used improved our understanding of the influence of those cognitive and non-cognitive factors on students' academic achievement.

### 5.2 Limitations of this research and recommendations for future research

The following limitations of the study were noted. First, the survey could have been conducted in two parts; however, the students needed more time than expected to complete it. Students took up to 15 min to complete the survey. Other researchers (notably, Gratiano and Palm, 2016), employed various techniques to reduce the survey completion time. They proposed a 5 min questionnaire to predict success and retention in first-year engineering students. The survey comprised three open-ended questions, and students were given 10 min to complete the survey. The participants were then interviewed for 30–40 min. However, it should be noted that the researcher conducting the survey reported that the (questionnaire and interview) technique was too lengthy for full participation by the students invited. They also noted that the surveys and interviews were somewhat invasive of student privacy because personal information was collected. Since the surveys and interviews were conducted by the professor, the students might not have felt comfortable providing accurate answers. In our study, the researcher conducting the surveys was a total stranger to the students; the researcher reported that the students were willing to answer all the survey questions and appeared both positive and assured during the surveys.

Second, the study only applies to first- and second-year students because the covered factors are those that are essential for student survival through the first few years of study. Third, if the study is conducted for second-year students, grades of general engineering courses can be included in the model for better estimation of students' GPAs. In addition, decision makers may need to consider the influence of other non-quantitative factors on academic achievement such as learning preferences and personality styles.

### 5.3 Implications for practice

Several implications can be made as below.

*5.3.1 A comprehensive tool for a systematically gathered information on student population.* The "Building Blocks of Excellence" instrument provides a readily accessible tool for teachers and administrators. College administrators and professors can use it to gain insight into their students' attitudes and perceptions toward engineering at different times of the year. They can use the results to compare differences, measure growth or progress over time and plan for remedial or enhancement programs. Faculty may use the results of such predictive model to guide curriculum development for courses offered to first-year engineering students (e.g. skill-building seminars).

*5.3.2 An efficient predictive modeling technique.* This study proposes an actionable, applicable and accurate technique for the prediction of student outcomes. Both methods can be used when there are a large number of variables/predictors. Since the differences in the  $R^2$  and MSE are not great, we recommend that either method can be used when there are a large number of variables/predictors.

However, when predictors are not related to each other, and if the number of predictors is not more than 20 we can use the best subset selection technique (James *et al.*, 2013). According to Blum *et al.* (2013), the best subset is conceptually simple but is difficult to manage for a large number of potential summary statistics. If we are more interested to

---

capture as much variance regarding the predicted outcome and are not interested in finding how the data can be grouped based on their linear combination, we may use the best subset. Moreover, no sample restrictions exist when using the best subset.

On the other hand, when we have a large number of variables and some of which are correlated. This correlation between variables brings about a redundancy in the information that can be gathered by the data set. For this reason, we may use PCA to transform the original variables to the linear combination of these variables which are independent. However, sample restrictions exist for the successful implementation of PCA, which can be checked through the sampling adequacy procedure.

Ridge regression is a robust technique that is not affected by outliers and does not require the dependent variable to follow a normal distribution as in the OLS method. It is both a simple and reliable technique which can be used as an alternative to standard multiple regression when the predictors are highly correlated. It also provides a more stable and interpretable model.

5.3.3 *A whole-institution approach to engineering students' success and retention.* According to Thomas *et al.* (2017), institutions of higher education that want to achieve excellence in learning and teaching, and improve student experiences, persistence and success should “adopt an evidence-informed, whole-institution approach to achieve change” (p. 28). Their report provided the basis for a successful initiative or approach aimed at improving student outcomes.

The most important implication is that establishing an engineering success lab/center which involves staff, professors, graduate students and undergraduate engineering students would be beneficial. This success center would aim to improve students' mindsets along with their social and academic integration. Some of the targeted initiatives that can be offered in this center may include the following.

5.3.3.1 Math helpdesk. An important service that can be offered through this center includes a math helpdesk, where a student could arrange for one-on-one or group tutoring.

5.3.3.2 Targeted workshops and training sessions. On the basis of determinants or critical-to-success factors, planned or targeted workshops on goal setting, study skills, social skills and time-management skills could be organized for students.

5.3.3.3 Internship and job shadowing. One of the best practices that can increase student motivation, commitment and self-efficacy is providing early opportunities for internships and job shadowing. These opportunities can be achieved through collaboration and partnership with the industry, which can help students develop their engineering identities.

5.3.3.4 Increase students' academic and social engagement opportunities. Providing students with opportunities to conduct research alongside faculty members can also greatly increase their academic self-efficacy and integration. By contrast, social integration can be achieved by engaging students in community-service events or programs. Such events can be led and managed by the students themselves. Service events may be included as course requirements and supervised by professors or offered as general volunteering hours. The supervision and support for these programs should be provided by the engineering success center.

5.3.3.5 Extensive support for low-performing students. In collaboration with the academic advising unit, this center should be responsible for the identification of low-performing students. Progress counseling is an essential practice to support low-performing students; progress counseling should include recommendations and strategies to enhance success (e.g. attending a series of skill-building workshops or summer enrichment programs).

Supporting engineering students in their early years of study can create a positive atmosphere for all students to excel. As indicated by Tinto's theory, when the student is successfully integrated socially and academically, he/she is more likely to persist and succeed. The advancement of our society, national security and quality of life is strongly

---

linked to scientific and technological innovation and development. Paying more attention to the field of engineering education is a key component toward a prosperous and innovative economy.

## 6. Conclusion

This study was based on well-established theoretical background and validated measures, and was designed to test the predictive capability of a combination of cognitive and non-cognitive factors. Our cognitive and non-cognitive factors were inspired by the KAB approach. We used a student's HSG to represent "knowledge," a number of motivational, attitudinal and self-efficacy factors to represent "attitudes," and a number of skill-related behaviors to represent "behavior." Among these, students' HSGs, academic mindset and academic self-discipline were strongly associated with success in engineering courses. The use of a gender/group diverse sample guaranteed that our predictive model accounts for all students including female and first-year students, who are the most prone to failure and withdrawal.

The results of this study suggest that vulnerable students can be identified as early as possible in the academic year. These students should be encouraged to engage more in their learning process. As educators, we should take greater responsibility for finding and utilizing all available information or opportunities available to enhance the education of our students. Carrying out such a measurement at the beginning of the college year can provide college staff and administrators with valuable insights into students' perception of their own skills and into their attitudes toward engineering. The college academic advisor can use this information to develop individualized plans to guide students on how to best work to enhance those factors that are critical to their success and retention. This early support would ensure a smooth transition of engineering undergraduates into the senior levels of study.

## References

- Ackerman, P.L., Kanfer, R. and Beier, M.E. (2013), "Trait complex, cognitive ability, and domain knowledge predictors of baccalaureate success, STEM persistence, and gender differences", *Journal of Educational Psychology*, Vol. 105 No. 3, pp. 911-927, doi: 10.1037/a0032338.
- Allen, J., Robbins, S.B., Casillas, A. and Oh, I.S. (2008), "Third-year college retention and transfer: effects of academic performance, motivation, and social connectedness", *Research in Higher Education*, Vol. 49 No. 7, pp. 647-664, doi: 10.1007/s11162-008-9098-3.
- Andrews, C.K. (2014), "Utilizing traditional cognitive measures of academic preparation to predict first-year science, technology, engineering, and mathematics (STEM) majors' success in math and science courses", available at: <http://digitalcommons.fiu.edu/cgi/viewcontent.cgi?article=2722&context=etd> (accessed June 30, 2017).
- Astin, A.W. (1993), *What Matters in College: Four Critical Years Revisited*, Jossey Bass Higher & Adult Education Series, San Francisco, CA.
- Astin, A.W. and Astin, H.S. (1992), "Undergraduate science education: the impact of different college environments on the educational pipeline in the sciences", final report, National Science Foundation, Washington, DC, available at: <http://files.eric.ed.gov/fulltext/ED362404.pdf> (accessed May 2, 2016).
- Bandura, A. (1994), "Self-efficacy", in Ramachandran, V.S. (Ed.), *Encyclopedia of Human Behavior*, Vol. 4, Academic Press, New York, NY, pp. 71-81.
- Bernadin, S.L. and McKendrick, W. (2015), "A study of non-academic factors contributing to the success of undergraduate engineering students", *ASEE Southeast Section Conference, Gainesville, FL, April*, available at: [www.softwareeducationssupport.com/ASEE%20SE%20Conference%20Proceedings/Conference%20Files/ASEE2015/Papers2015/103.pdf](http://www.softwareeducationssupport.com/ASEE%20SE%20Conference%20Proceedings/Conference%20Files/ASEE2015/Papers2015/103.pdf) (accessed February 26, 2016).

- Bernold, L.E., Spurlin, J.E. and Anson, C.M. (2007), "Understanding our students: a longitudinal-study of success and failure in engineering with implications for increased retention", *Journal of Engineering Education*, Vol. 96 No. 3, pp. 263-274, doi: 10.1002/j.2168-9830.2007.tb00935.x.
- Besterfield-Sacre, M., Atman, C.J. and Shuman, L.J. (1997), "Characteristics of freshman engineering students: models for determining student attrition in engineering", *Journal of Engineering Education*, Vol. 86 No. 2, pp. 139-149, doi: 10.1002/j.2168-9830.1997.tb00277.x.
- Besterfield-Sacre, M., Atman, C.J. and Shuman, L.J. (1998), "Engineering student attitudes assessment", *Journal of Engineering Education*, Vol. 87 No. 2, pp. 133-141, doi: 10.1002/j.2168-9830.1998.tb00333.x.
- Besterfield-Sacre, M., Moreno, M., Shuman, L.J. and Atman, C.J. (1999), "Comparing entering freshman engineers: institutional differences in student attitudes", *Age*, Vol. 4 No. 4, pp. 133.1-133.12.
- Blum, M.G., Nunes, M.A., Prangle, D. and Sisson, S.A. (2013), "A comparative review of dimension reduction methods in approximate Bayesian computation", *Statistical Science*, Vol. 28 No. 2, pp. 189-208, doi: 10.1214/12-STS406.
- Britton, B.K. and Tesser, A. (1991), "Effects of time-management practices on college grades", *Journal of Educational Psychology*, Vol. 83 No. 3, pp. 405-410, doi: 10.1037/0022-0663.83.3.405.
- Burtner, J. (2004), "Critical-to-quality factors associated with engineering student persistence: the influence of freshman attitudes", *34th ASEE/ISEE Frontiers in Education Conference, Savannah, GA, October 20-30*, doi: 10.1109/FIE.2004.1408601.
- Burtner, J. (2005), "The use of discriminant analysis to investigate the influence of non-cognitive factors on engineering school persistence", *Journal of Engineering Education*, Vol. 94 No. 3, pp. 335-338, doi: 10.1002/j.2168-9830.2005.tb00858.x.
- Chatterjee, S. and Hadi, A.S. (2015), *Regression Analysis by Example*, John Wiley & Sons, New York, NY.
- Deci, E.E. and Ryan, R.M. (1985), *Intrinsic Motivation and Self-Determination in Human Behavior*, Plenum Press, New York, NY.
- Deci, E.L. and Ryan, R.M. (1991), "A motivational approach to self: integration in personality", *Perspectives on Motivation*, Vol. 38 No. 1, pp. 237-288.
- De Winter, J.C.F. and Dodou, D. (2011), "Predicting academic performance in engineering using high school exam scores", *International Journal of Engineering Education*, Vol. 27 No. 6, pp. 1343-1351, available at: [www.ece.uvic.ca/~rexlei86/SPP/otherswork/19\\_ijee2519ns.pdf](http://www.ece.uvic.ca/~rexlei86/SPP/otherswork/19_ijee2519ns.pdf)
- Duran, R.L. (1983), "Communicative adaptability: a measure of social communicative competence", *Communication Quarterly*, Vol. 31 No. 4, pp. 320-326, doi: 10.1080/01463378309369521.
- Eccles, J.S. and Wigfield, A. (2002), "Motivational beliefs, values, and goals", *Annual Review of Psychology*, Vol. 53 No. 1, pp. 109-132, doi: 10.1146/annurev.psych.53.1.109.
- El-Dereny, M. and Rashwan, N.I. (2011), "Solving multicollinearity problem using ridge regression models", *International Journal of Contemporary Mathematical Sciences*, Vol. 6 No. 12, pp. 585-600, available at: [www.m-hikari.com/ijcms-2011/9-12-2011/rashwanIJCMS9-12-2011.pdf](http://www.m-hikari.com/ijcms-2011/9-12-2011/rashwanIJCMS9-12-2011.pdf)
- Erickson, G.M. (1981), "Using ridge regression to estimate directly lagged effects in marketing", *Journal of the American Statistical Association*, Vol. 76 No. 376, pp. 766-773, doi: 10.1080/01621459.1981.10477719.
- Farruggia, S.P., Han, C.W., Watson, L., Moss, T.P. and Bottoms, B.L. (2018), "Noncognitive factors and college student success", *Journal of College Student Retention: Research, Theory and Practice*, Vol. 20 No. 3, pp. 308-327, doi: 10.1177/1521025116666539.
- French, B.F., Immekus, J.C. and Oakes, W. (2003), "A structural model of engineering students success and persistence", *33rd Annual Frontiers in Education*, Vol. 1, IEEE, Boulder and Westminster, CO, November 5-8, pp. T2A-19, doi: 10.1109/FIE.2003.1263286.
- French, B.F., Immekus, J.C. and Oakes, W.C. (2005), "An examination of indicators of engineering students' success and persistence", *Journal of Engineering Education*, Vol. 94 No. 4, pp. 419-425, doi: 10.1002/j.2168-9830.2005.tb00869.x.

- Geiser, S. and Veronica Santelices, M. (2007), "Validity of high-school grades in predicting student success beyond the freshman year: high-school record vs. standardized tests as indicators of four-year college outcomes", Research and Occasional Papers Series, University of California, Berkeley, available at: <https://escholarship.org/uc/item/7306z0zf> (accessed June 1, 2017).
- Gore, P.A. (2006), "Predicting the performance and persistence of first-year college students: the role of non-cognitive variables", *25th Annual Conference of the First-Year Experience, Atlanta, GA, February 24-28*.
- Gratiano, S.M. and Palm, W.J. (2016), "Can a five minute, three question survey foretell first-year engineering student performance and retention?", *Proceedings of the 123rd ASEE Annual Conference and Exposition, New Orleans, LA, June*.
- Gray, G., Mcguinness, C. and Owende, P. (2016), "Non-cognitive factors of learning as early indicators of students at-risk of failing in tertiary education", in Khine, M.S. and Areepattamannil, S. (Eds), *Non-cognitive Skills and Factors in Educational Attainment. Contemporary Approaches to Research in learning Innovations*, Sense Publishers, Rotterdam, pp. 199-237.
- Green, P.J. and Silverman, B.W. (1993), *Nonparametric Regression and Generalized Linear Models: A Roughness Penalty Approach*, CRC Press, Washington, DC.
- Guadagnoli, E. and Velicer, W.F. (1988), "Relation to sample size to the stability of component patterns", *Psychological Bulletin*, Vol. 103 No. 2, pp. 265-275, doi: 10.1037/0033-2909.103.2.265.
- Guay, F., Vallerand, R.J. and Blanchard, C. (2000), "On the assessment of situational intrinsic and extrinsic motivation: the Situational Motivation Scale (SIMS)", *Motivation and Emotion*, Vol. 24 No. 3, pp. 175-213, doi: 10.1023/A:1005614228250.
- Hall, C.W., Kauffmann, P.J., Wuensch, K.L., Swart, W.E., DeUrquidi, K.A., Griffin, O.H. and Duncan, C.S. (2015), "Aptitude and personality traits in retention of engineering students", *Journal of Engineering Education*, Vol. 104 No. 2, pp. 167-188, doi: 10.1002/jee.20072.
- Hiss, W.C. and Franks, V.W. (2014), "Defining promise: optional standardized testing policies in American college and university admissions", The National Association for College Admission Counseling, Minot, ME.
- Hoerl, A.E. and Kennard, R.W. (1970), "Ridge regression: biased estimation for nonorthogonal problems", *Technometrics*, Vol. 12 No. 1, pp. 55-67.
- Huang, D., Guan, P., Guo, J., Wang, P. and Zhou, B. (2008), "Investigating the effects of climate variations on bacillary dysentery incidence in northeast China using ridge regression and hierarchical cluster analysis", *BMC Infectious Diseases*, Vol. 8 No. 1, pp. 8-130, doi: 10.1186/1471-2334-8-130.
- Huang, S. (2011), "Predictive modeling and analysis of student academic performance in an engineering dynamics course", Utah State University, Logan, UT, available at: <http://digitalcommons.usu.edu/etd/1086/> (accessed November 21, 2017).
- James, G., Witten, D., Hastie, T. and Tibshirani, R. (2013), *An Introduction to Statistical Learning*, Vol. 112, Springer, New York, NY.
- Jiang, X. and Freeman, S. (2011), "An analysis of the effect of cognitive factors on students' attritions in engineering: a literature review", *Proceedings of STEM Agenda Conference on the American Society for Quality, WI, July 19-20*, available at: <http://rube.asq.org/edu/2011/07/engineering/an-analysis-of-the-effect-of-cognitive-factors-on-students-attrition-in-engineering-a-literature-paper.pdf> (accessed April 16, 2016).
- Jin, Q., Imbrie, P.K., Lin, J.J. and Chen, X. (2011), "A multi-outcome hybrid model for predicting student success in engineering", American Society for Engineering Education, Washington, DC, available at: [www.asee.org/file\\_server/papers/attachment/file/0001/1611/Mar14\\_NNs.pdf](http://www.asee.org/file_server/papers/attachment/file/0001/1611/Mar14_NNs.pdf) (accessed November 1, 2017).
- Jørgensen, U. and Valderrama, A. (2016), "The politics of engineering professionalism and education", in Jørgensen, U. and Brodersen, S. (Eds), *Engineering Professionalism*, Sense Publishers, Rotterdam, pp. 283-309, available at: [https://doi.org/10.1007/978-94-6300-752-8\\_14](https://doi.org/10.1007/978-94-6300-752-8_14) (accessed March 10, 2017).

- Kauffmann, P., Abdel-Salam, T. and Garner, J. (2007), "Predictors of success in the first two years: a tool for Retention", *Annual Conference of American Society for Engineering Education, Pacific Southeast, June*.
- Kim, J.O. and Mueller, C.W. (1978), *Factor Analysis: Statistical Methods and Practical Issues*, Vol. 14, Sage, Beverly Hills, CA.
- Le, H., Casillas, A., Robbins, S.B. and Langley, R. (2005), "Motivational and skills, social, and self-management predictors of college outcomes: constructing the student readiness inventory", *Educational and Psychological Measurement*, Vol. 65 No. 3, pp. 482-508, doi: 10.1177/0013164404272493.
- Lin, J.J. and Imbrie, P.K. (2014), "Modeling retention and graduation of engineering students of difference sexes", *6th First Year Engineering Experience Conference, TX, August 7-8*, available at: <http://fyee.org/fyee2014/papers/1003.pdf> (accessed January 20, 2016).
- Lin, J.J., Malyschiff, A. and Imbrie, P.K. (2008), "Artificial intelligence methods to forecast engineering students' retention based on cognitive and non cognitive factors", available at: <https://peer.asee.org/4315> (accessed October 7, 2016).
- Lotkowski, V.A., Robbins, S.B. and Noeth, R.J. (2004), "The role of academic and non-academic factors in improving college retention", ACT policy report, Iowa City, IA.
- May, D.K. (2009), "Mathematics self-efficacy and anxiety questionnaire", doctoral dissertation, University of Georgia, Athens, GA.
- National Academy of Engineering (2004), *The Engineer of 2020: Visions of Engineering in the New Century*, The National Academies Press, Washington, DC, doi: 10.17226/10999.
- National Science Foundation (2007), *Moving Forward to Improve Engineering Education*, National Science Foundation, Washington, DC, available at: [www.nsf.gov/pubs/2007/nsb07122/nsb07122.pdf](http://www.nsf.gov/pubs/2007/nsb07122/nsb07122.pdf) (accessed July 12, 2017).
- Pintrich, P.R. and De Groot, E.V. (1990), "Motivational and self-regulated learning component of classroom academic performance", *Journal of Educational Psychology*, Vol. 82 No. 1, pp. 33-40, doi: 10.1037/0022-0663.82.1.33.
- Poropat, A.E. (2009), "A meta-analysis of the five-factor model of personality and academic performance", *Psychological Bulletin*, Vol. 135 No. 2, pp. 322-338, doi: 10.1037/a0014996.
- Reid, K.J. (2009), "Development of the student attitudinal success instrument: assessment of first year engineering students including differences by gender", ProQuest Dissertations Publishing, Purdue University, West Lafayette, IN.
- Richardson, M., Abraham, C. and Bond, R. (2012), "Psychological correlates of university students' academic performance: a systematic review and meta-analysis", *Psychological Bulletin*, Vol. 138 No. 2, pp. 353-387, doi: 10.1037/a0026838.
- Robbins, S.B., Lauver, K., Le, H., Davis, D., Langley, R. and Carlstrom, A. (2004), "Do psychosocial and study skill factors predict college outcomes? A meta-analysis", *Psychological Bulletin*, Vol. 130 No. 2, pp. 261-288, doi: 10.1037/0033-2909.130.2.261.
- Rubin, R.B. and Martin, M.M. (1994), "Development of a measure of interpersonal communication competence", *Communication Research Reports*, Vol. 11 No. 1, pp. 33-44, doi: 10.1080/08824099409359938.
- Saltonstall, M. (2013), "Predicting college success: achievement, demographic, and psychosocial predictors of first-semester college grade point average", doctoral dissertation, Northern Arizona University, Flagstaff, AZ, available at: <https://search.proquest.com/docview/1492736859?pq-origsite=gscholar> (accessed September 9, 2016).
- Sax, L.J. (2008), *The Gender Gap in College: Maximizing the Developmental Potential of Women and Men*, Jossey-Bass, San Francisco, CA.
- Schiavone, P. (2001), *Engineering Success*, Prentice Hall, Upper saddle River, NJ.
- Schrader, P.G. and Lawless, K.A. (2004), "The knowledge, attitudes, and behaviors approach how to evaluate performance and learning in complex environments", *Performance Improvement*, Vol. 43 No. 9, pp. 8-15, doi: 10.1002/pfi.4140430905.

- Sheppard, S., Gilmartin, S., Chen, H.L., Donaldson, K., Lichtenstein, G., Eris, O., Lande, M. and Toye, G. (2010), "Exploring the engineering student experience: findings from the academic pathways of people learning engineering survey (APPLES)", Center for the Advancement of Engineering Education, Seattle, WA, available at: <http://files.eric.ed.gov/fulltext/ED540124.pdf> (accessed June 17, 2015).
- Stankov, L. and Lee, J. (2014), "Quest for the best non-cognitive predictor of academic achievement", *Educational Psychology*, Vol. 34 No. 1, pp. 1-8, doi: 10.1080/01443410.2013.858908.
- Thomas, L., Hill, M., O'Mahony, J. and Yorke, M. (2017), "Supporting student success: strategies for institutional change: what works? Student retention and success programme", Document by UK Higher Education Academy, London.
- Ting, S.M.R. and Man, R. (2001), "Predicting academic success of first-year engineering students from standardized test scores and psychosocial variables", *International Journal of Engineering Education*, Vol. 17 No. 1, pp. 75-80, available at: [www.ijee.ie/articles/Vol17-1/Ijee1168.pdf](http://www.ijee.ie/articles/Vol17-1/Ijee1168.pdf)
- Tinto, V. (1993), *Leaving College: Rethinking the Causes and Cures of Student Attrition*, 2nd ed., University of Chicago Press, Chicago, IL.
- Trapmann, S., Hell, B., Hirn, J.O.W. and Schuler, H. (2007), "Meta-analysis of the relationship between the big five and academic success at university", *Zeitschrift für Psychologie/Journal of Psychology*, Vol. 215 No. 2, pp. 132-151, available at: <http://econtent.hogrefe.com/doi/abs/10.1027/0044-3409.215.2.132>
- Veenstra, C.P. (2008), "Modeling freshman engineering success", doctoral dissertation, University of Michigan, Ann Arbor, MI, available from: <http://hdl.handle.net/2027.42/58391> (accessed November 1, 2017).
- Veenstra, C.P., Dey, E.L. and Herrin, G.D. (2009), "A model for freshman engineering retention", *Advances in Engineering Education*, Vol. 1 No. 3, pp. 1-31, available at: <http://files.eric.ed.gov/fulltext/EJ1076050.pdf>
- Virtanen, P., Nevgi, A. and Niemi, H. (2013), "Self-regulation in higher education: students' motivational, regulational and learning strategies, and their relationships to study success", *Studies for the Learning Society*, Vol. 3 Nos 1-2, pp. 20-34, doi: 10.2478/sls-2013-0004.
- Vrugt, A. and Oort, F.J. (2008), "Metacognition, achievement goals, study strategies and academic achievement: pathways to achievement", *Metacognition and Learning*, Vol. 3 No. 2, pp. 123-146, doi: 10.1007/s11409-008-9022-4.
- Wall, K. (2010), *Engineering: Issues, Challenges and Opportunities for Development*, UNESCO, Paris, available at: <http://unesdoc.unesco.org/images/0018/001897/189753e.pdf> (accessed November 12, 2017).
- Walton, S.D. (2010), "Exploring the relationship between resilience and learning styles as predictors of academic persistence in engineering", doctoral dissertation, Texas A&M University, TX, available at: <http://oaktrust.library.tamu.edu/bitstream/handle/1969.1/ETD-TAMU-2010-12-8954/WALTON-DISSERTATION.pdf> (accessed August 5, 2016).
- Whitaker, X.L. (2014), "The role of psychosocial factors in the academic success of engineering majors", doctoral dissertation, Azusa Pacific University, Azusa, CA.
- Whittaker, J.C., Thompson, R. and Denham, M.C. (2000), "Marker-assisted selection using ridge regression", *Genetics Research*, Vol. 75 No. 2, pp. 249-252.

### Further reading

James, H. (1937), *The Ambassadors*, Scribner, New York, NY.

### About the authors

Dr Bothaina A. Al-Sheeb holds Bachelor Degree in English language and literature from Qatar University. She gained her Masters in education leadership in 2009 from the same institution. She has worked in local Qatari schools for several years. In 2010, she joined the Academic Bridge Program as Business Coordinator in Qatar Foundation. She is recent PhD holder in engineering management,

from Qatar University, Qatar, 2018. Her main research interests are: engineering education, first-year experience, modeling student performance and retention. Dr Bothaina A. Al-Sheeb is the corresponding author and can be contacted at: 200669067@qu.edu.qa

Dr A.M. Hamouda is currently the Associate Dean for Research and Graduate Studies, College of Engineering, Qatar University. Prior to this appointment, Dr Hamouda held the position of Chairman of Mechanical and Industrial Engineering Department, College of Engineering, Qatar University. Dr Hamouda held joint appointments in University Putra Malaysia, Faculty of Engineering as well as the Institute of Advanced Technology where he was the Founding Program Manager in computational Science and Engineering. Dr Hamouda is Corresponding Member for the American Academy of Mechanics (AAM), Member of the American Society of Mechanical Engineering (ASME). He is also Senior Member of Institute of Industrial Engineering (IIE), USA and Member of the Institute of Highway Transportation, UK.

Dr Galal M. Abdella is currently Assistant Professor. He holds PhD in quality management and control, from Wayne State University, USA, 2011. His research areas include: high-dimensional process monitoring and control and modeling and simulating real-world systems.