Using Regression to Establish Weights for a Set of Composite Equations through a Numerical Analysis Approach: A Case of Admission Criteria to a College

[1]Ramzi Naim Nasser and [2]Viviane Naimy
[1]Department of Educational Sciences, College of Education, Qatar University, Qatar
[2]Department of Finance, Faculty of Business and Economics, Notre Dame University, Lebanon

Abstract: Problem statement: Mathematically little is known of college admission criteria as in school grade point average, admission test scores or rank in class and weighting of the criteria into a composite equation. **Approach:** This study presented a method to obtain weights on "composite admission" equation. The method uses an iterative procedure to build a prediction equation for an optimal weighted admission composite score. The three-predictor variables, high school average, entrance exam scores and rank in class, were regressed on college Grade Point Average (GPA). The weights for the composite equation were determined through regression coefficients and numerical approach that correlate the composite score with college GPA. **Results:** A set of composite equations were determined with the weights on each criteria in a composite equation. **Conclusion:** This study detailed a substantiated algorithm and based on an optimal composite score, comes out with an original and unique structured composite score equation for admissions, which can be used by admission officers at colleges and universities.

Key words: Composite equations, weights, standardized regression coefficients, iterative numerical approach

INTRODUCTION

Institutions of higher education vary considerably in the degree to which they are selective in admissions. Typically, students may get rejected or accepted to a major of their choice based on a set of academic criteria that a university establishes and uses. This variety in selectivity, perhaps, is one reason that admissions practices are not generally well understood by either the applicants or their parents.

Research studies have not tackled the development of a logical, mathematical, interpretive and algorithmic model in the selection and weighting of admission criteria to colleges and universities. The use of a two-variable criteria for admissions to predict admission (Hu, 2002) as in High School Grade Point Average (HSGPA) and standardized tests, such as the Scholastic Achievement Test (SAT) that predict (directly or indirectly) an applicant's probability of academic success in the first year of college. However, few studies have emerged where the index on each criterion was not validated through an empirical numerical and computational approach. In this study, we fill this gap by proposing a way to establish weights for admissions and validate these weights using numerical methods that build a composite admission score. A common procedure for admission decisions is to select a number of predictors for general performance at college (videlicet, GPA), assign subjective weights to criteria that predict performance and then build a composite score (McCormick and Ilgen, 1980). To validate the composite equations, the scores on the equations will be correlated with the third semester GPA. The College Board and the Trends in College Admission Report (Breland *et al*., 2002) show an overwhelming number of US universities requiring HSGPA, standardized tests, or Rank In Class (RIC) for admissions to college.

Since there is a complete absence of a unified and standardized set of criteria for admissions in US-style universities around the world, an operational structured approach, describing and justifying the selection and weighting process of admission criteria is fundamental and informative for parents and students. The main goal of this research is to find a statistically valid, mathematically logical and applicable admission method aiming at establishing appropriate weights for a university admissions criterion.

Corresponding Author: Ramzi Nasser, Department of Educational Sciences, Center of Educational Development and Research, Qatar University, Qatar

This study develops a model to weight the admission criteria to a college: High School Average (HSA), Entrance Exam Scores (EES) or SAT and RIC. The criteria and weights are summarized in a composite equation. The model is algorithmic and applies a regression analysis, which then correlates the regression line with ordinal weighted criteria on a predicted college GPA (Maximal correlations between the average weighted criteria for the composite and college GPA dictates the weight for each criteria). Based on the regression, a decision-tree algorithm iterates a correlation between the observed college GPA and the result of the ordinal-weighted composite equation that predicts college GPA. The repetitive calculations of an ordinal weighted criteria and repetitive correlation is what we call as the numerical analysis that weights the admission criteria, which in turn is iterated by decreasing or increasing the coefficients of each criteria. The selection of the optimal composite equation has different average-weighted criteria into a composite score, achieved through the correlation between a composite score and college GPA.

Review of literature: There is a general trend among universities and colleges to use a reliable basic skills test (home-made within a college or a university), HSGPA and the SAT scores, for admittance into college (Breland *et al.*, 2002). Some leeway in the use of admission criteria in US universities suggests a greater support for HSGPA as the predominant criteria for admittance. The most frequent requirement, for both public and private institutions in the US, is HSGPA. The trends report indicates that a little more than one-half of all four-year institutions in 2000 required a minimum HSGPA and a standardized admission tests like the SAT, underlining the diminishing level and use of RIC for admittance (only about ¼ of the four-year institutions reported minimum standards for high school rank).

A number of methods have been established to determine the weights for an admission index. Earlier studies show that both HSGPA and the standardized test scores are the strongest predictors for institutional selectively of students (Weizman, 1982). Even, with many criteria empirically tested, high school average is a slightly better predictor of first-year college GPAs of 2.00 or higher than were admission college test scores, with HSGPA (Noble and Sawyer, 2000; 2002). HSGPA showing a significant prediction power for institutional selectivity of students (Noble and Sawyer, 2002) a primary predictor to student success in graduate work. Generally, maximal correlations between the average weighted criteria for the

composite and college GPA dictates the weight for each criterion) (Paolillo, 1982; Youngblood and Martin, 1982). A number of other studies have also confirmed prediction power and use of these variables such as high school rank, standardized test score (ACT or SAT scores) and RIC to predict college performance (Price and Kim, 1976; Mouw and Khanna, 1993; Beecher and Fischer, 1999; Noble and Sawyer, 2000; 2002; Xiao, 2002).

A variety of studies have provided evidence and criteria for college success (Paolillo, 1982; Sisson and Dizney, 1980), generally showing a correlation between university input measures such as standardized test scores, national exams, SAT and HSGPA on college GPA (Eno *et al.*, 1999; Noble and Sawyer, 2000).

The use of the regression method is probably one of the most widely used methods for establishing balanced weights for admission criteria (McCormick and Ilgen, 1980). Other compatible weighting methods have also been compared (Fralicx and Raju, 1982) indicating the similarities of these methods. While a logistic regression analysis to predict binomial or discrete outcomes as a valid admission index to college success (Xiao, 2002) However, not one study has approached the weighting of criteria using a conservative, operative and logical numerical method to validate the weights for a set of admission criteria. Hence, a composite admission score based on regression provides a balanced weight for each criterion and a standard and indicator to whether students get admitted or not (Talley, 1989; Talley and Mohr, 1991).

MATERIALS AND METHODS

The data: Data from 2001-2006 of all undergraduate enrollees in a private university. The study used the base-semester data starting in the fall of 2001. All identifiers as names of students were removed from the data set. The data was accrued from the student record's system. The data included high school average of the last 2 years of secondary School Entrance Exams (EES) or SAT I-Reasoning scores and RIC. The third semester GPA was used as a dependent variable. For example, data as high school average, EES and RIC from fall of 2001 were entered in the regression to predict college GPA (third-semester GPA).

To counter for irregularity in grading scales, high school grades were standardized for each school and a standard score (z) calculated using the mean and the standard deviation of the same school. The z was then converted to nonstandard normal distribution and then to percentage score.

The method in this study used regression to predict the third semester after enrollment. Two sets of regression equations were used, the first regression equation included two variables: high school average and EES. The use of the two-variable predictors is because student's RIC was not available; as many schools do not maintain a RIC or those who enrolled with the International Baccalaureate scores had no RIC. For students who had RIC, a second regression was used and it included three predictor variables: High school average, EES (or the SAT-I) and RIC. Regression coefficients were calculated and standardized coefficient ratios determined the weighting order of the admission criteria (i.e., high school average, EES and RIC).

RESULTS

Regression using high school average and entrance exam scores: All grades whether high school average, EES or RIC were all converted to percentage scores. In the first regression, high school average and EES were regressed on the third semester GPA. The first regression equation was based on 2947 data points. The standardized coefficients of high school average and EES generated by the regression analysis are shown in Table 1. The standardized betas of the regression were used to obtain a ratio referred to as the Standardized Regression Coefficient Ratio (SRCR). The ratio of 1.55: 1 (the SRCR: 0.34/0.22 = 1.55, Table 1) for high school average to EES indicates, at some permutable level, a higher weight for high school average to EES (or SAT I), which indicates that every unit, change in EES gives 55% greater change in high school average.

The regression equation was set-up such that an ordinal-weighted scheme was applied on each variable of high school average and EES, by increasing the weight on one and decreasing the weight on the other and its converse. A weighted change on each criterion of the predicted equation was then used. The weighted increase/decrease is generally dictated by the SRCR value, thus, by using a maximum weight increase of

55% higher per unit change in EES. An iterated predicted GPA value was calculated using the regression equation. Thus, the iterations defined by the SRCR are indexed by i (i = 1–22). For each iteration, a 5% weight change was used for the changed criterion. The iteration process can be expressed in Eq. 1:

$$Y_i = \varepsilon + (1+I_i) X_{ij}\beta_1 + (1-I_i) X_{ij} \beta_2 \qquad (1)$$

I_i = 0.05-0.55, at an increment of 0.05
$(-)I_i$ = (-) 0.05 to (-) 0.55, at an increment of (-) 0.05
i = 1- 22
j = 1-2947

$$PCGPA = 1.403 + (1+I) \; HSA \times 0.012 + (1-I.)$$
$$(EES) \times 0.009 \qquad (2)$$

PCGPA = Predicted College GPA
I = Increment
HAS = High School Average

For each iteration, a correlation r_i was calculated for the predicted GPA (Y_i), with the actual third semester GPA. The highest correlation appeared for a weighted coefficient of I = 0.1 and I = -0.1, a 10% increase/decrease in EES and its converse. This gave a tolerance range of 10 to -10% weight increase/decrease used to determine the weighted criteria for the actual values of high school average and EES on the composite Eq. 2.

Calculating the weights: A composite equation was constructed based on the criteria of high school average and EES. Through SRCR, a directional increase in weight was allocated for high school average and a decrease in EES. The composite equation was reiterated 10 times in a positive directional weight increase of high school average by a maximum tolerance of 10% and incremented by 1% and a decrease for the weight of EES by 1% reaching a maximum tolerance of 10% decrease.

Table 1: Regression results using high school average and EES

	Unstandardized coefficient ß	Standard error	Standardized coefficient ß	t	Significance
Constant	1.4030	0.063		22.297	0.000
High school average	0.0120	0.001	0.34	12.72	0.000
Entrance exam scores	0.0089	0.001	0.22	8.18	0.000

Table 2: Regression results using high school average, entrance exam scores and student rank

	Unstandardized coefficient ß	Standard error	Standardized coefficient ß	t	Significance
Constant	1.630	0.053		30.733	0.000
RIC	0.003	0.001	0.126	3.733	0.000
Math entrance exam	0.009	0.001	0.239	9.265	0.000
High school average	0.007	0.001	0.281	8.197	0.000

Considering that each variable would weight equally and add up to a composite of 100%, these criteria were in percentage units and then each would have a 50% chance of the total score (Eq. 3). Therefore, the composite score formula was weighted such that high school average had a higher weight than EES, where a greater proportion was given to high school average (as shown through SRCR). At each iteration-an increase in high school average (above the 50%) and decrease in EES (below the 50%)-a correlation was calculated between the composite score and the third-semester GPA. The highest correlation r = 0.48, p<0.0001 appeared for an increase of 10% to a cumulative of 60% weight for high school average and a decrease by 10% to a reduced cumulative of 40% for EES. The results merely validate the weighted predictor regression of college GPA. The final weighted score structure is shown in Eq. 4:

$$CS = ((1+I_i) \times + (1-I_i) \times EES_i)/2 \qquad (3)$$

CS = Composite Score

for I = 0.01 to 0.1 at increment of 0.01
i = 1 to 2947

$$CS = (0.55) \times HSA_i + (0.45) \times EES_i \qquad (4)$$

Regression 2: Using high school average, entrance exam scores and student's rank: In the previous algorithm, two variable criteria were used to determine a composite score. When there are more than two criteria, as high school average, EES and RIC, the algorithm goes through a decision-tree heuristic. The criteria as high school average, EES and RIC were regressed on the third-semester GPA. The analysis was based on 1207 data points.

The standardized coefficients of high school average, EES and students' RIC results are obtained from the regression analysis presented in Table 2 and its application in Eq. 5:

$$PCGPA = 1.630+0.007 \times HSA+0.009 \times \\ (EES) + 0.003 \times RIC \qquad (5)$$

A similar methodological approach to the two-variable criteria was used with the three variable criteria. However, a step-wise weighting method based on a decision-tree algorithm produced two SRCR corresponding to two tolerance levels. The SRCR showed that the highest standardized regression beta was for high school average followed by EES and lastly, by RIC. Thus, two comparisons were made: the first between high school average and EES and second between EES and RIC. There was no need to calculate

a third SRCR between high school average and RIC because a transitive relation can be implied. The weight ratio of high school average and EES was at 1.1757:1, which indicates that high school average has an 18% higher weight for a one-unit increase in EES scores. The weight ratio of EES to RIC was 1.89:1.

The weights were constrained through an increase/decrease dictated by the SRCR; thus, by holding RIC constant and using a 5% (i.e., 0.05) on the increased weight for average high school grade and simultaneous decrease in EES of 5% and its converse, a correlation was calculated between the predicted, weighted college GPA with actual third-semester GPA:

$$Yi = \epsilon + (1+I_i) X_{ij}B_1 + (1+Ii) XijB_2 + X_{ij}B_3 \qquad (6)$$

for I = 0.05, 0.1, 0.15, 0.20, -0.05, -0.1, -0.15 and -0.2
i = 1 to 8
j = 1 to 1207

$$PCGPA = 1.603 +(1+I) \times HSA \times (0.007) + (1-I)EES \\ \times (0.009)+RIC \times (0.003) \qquad (7)$$

PCGPA = Predicted College GPA.

The predicted college GPA for the first regression equation (Eq. 6) showed the highest correlation for a tolerance level for an increment between -0.1 and 0.1, (i.e., 10% decrease/increase). A second regression held high school average constant and EES and RIC weighted by a 0.1 increase/decrease (Eq. 8 and 9). At each increment increasing EES by a 0.1 increment and decreasing by 0.1 for RIC, a correlation was calculated between the predicted college GPA and actual GPA:

$$Yi = \epsilon + X_{ij}B_1 + (1+I)_i X_{ij}B_2 + (1-I)_i X_{ij}B_3 \qquad (8)$$

for I = 0.1, 0.20, 0.3, 0.4, 0.5, 0.6, 0.7, 0.8, 0.9
and for -I
i = 1 to18
j = 1 to 1207

$$PCGPA = 1.603+HSA \times (0.007)+(1-I) \times EES \\ \times (0.009)+(1+I) \times RIC \times (.03) \qquad (9)$$

The results of the second regression showed that at a 10% increase/decrease for EES and RIC respectively gave the highest correlation coefficient of the predicted college GPA and the actual third-semester college GPA. The composite equation was formulated based on the decision-tree algorithm. The weights for the pre-selected variable with the highest SRCR showing an increase in high school average by a tolerance level of 10% that would provide the highest correlation with the third-semester GPA and hence, high school average having a 10% increase in weight. The other two variables would have a 10% decrease in weight to

compensate for the increase in high school average. Because of the two-step increase/decrease in the weighted ratio approach, the variable that has the highest standardized regression coefficient, viz., high school average would have the highest weight. However, it could be implied that each variable (high school average, EES and RIC) has one-third of the total score. This would not hold in the weighting scheme because of the greater weight ratio given to high school average when compared with EES and RIC. In effect, the composite formula would be distributed equally between high school average, on the one hand and EES and RIC, on the other. Thus, each would have a 0.5 chance for a full score. This allocation is based on the SRCR where high school average produced a 110% cumulative increase over EES and RIC combined. Thus, allocating 50% of the weight to high school and 50% for EES and RIC combined gives a substantive conservative approach. Therefore, a 10% increase in high school average would require a 10% decrease in EES with RIC combined.

We introduced a range of reduced weights for the two combined variables EES and RIC. The increment change for the two combined variables ranged between 0.9 and -0.9 with 0.1 incremented runs as to correlate with third-semester GPA; this can be expressed in the following Eq. 10:

$$PCS = (1.1 \times HSA)/2 + (EES \times (0.9-I) + RIC \times (I))/2 \quad (10)$$

PCS = Predicted Composite Score.

The composite Eq. 10 had the weights decrease then increase on EES and RIC (ordered according to SRCR) from a value of -0.1 to -0.8 and from 0.-0.8 respectively; at an increment of 0.1. At each increment a claculated predicted composite score, was correlated with third-semester GPA. The highest correlation appeared for a 0.1 increment (10%), which decreases the EES by 0.8/2 and RIC by 0.1/2. The best possible weights for a valid, appropriate and empirically substantiated weight according to the final output of the three iterations could be calculated in the following manner:

- Weight for high school average= $((1.1)/2) \times 100 = 55\%$
- Weight for EES = $((0.9-0.1)/2) \times 100 = 40\%$
- Weight for Rank in Class = $(0.1/2) \times 100 = 5\%$

The final composite equation structure and weighted coefficients is expressed in 11:

$$CS = (0.55) HAS + (0.4)EES + (0.05)RIC \quad (11)$$

DISCUSSION

This study illustrates the development of a logical, interpretive and substantive academic admission composite score. The composite score is simply a weighted index of high school average, EES or SAT and RIC. The study validates the weighted composite score by correlating the score with the third-semester college GPA. The algorithm uses regression, which typically is used to validate predictive criteria for college GPA. Once the regression coefficients are calculated, they create the course for weight tolerance levels and provide the directional weight increase/decrease to the criteria used in the composite equation.

It is shown through numerical methods that the correlation with college GPA of a composite equation constitutes an evidence of an objective measure and calculation. The method can be seen as a model for which other variables can be used as criteria for admission. The method is very conservative in that it uses the regression equation and establishes the weights based on the regression coefficients to obtain a tolerance level as a best predictor for GPA. The predictor variables accurately use the SRCR as the guide for weighted limits. Homogeneously distributed weights on the composite equations were altered through a decision-tree approach, which alternates the weights for each variable sequentially with the weighted change between a maximum of two variables (criteria) at a time, until all possible combinations are exhausted. The composite score equation is then weighted with respect to the tolerance levels. The highest correlation between predicted college GPA and actual college GPA provides the optimal weighted criteria on a composite equation. This validation overlaps with the predicted regression equation with its weighted criteria. The significance of this method is that it can effectively predict college success and thus help admission officers set the criteria for admission to a college. The composite score can also define and can be used for cut-off scores based on admission policy. We would also show in future work that admission cut-offs could be empirically validated through numerical methods (iterative procedures) and thus, not altogether a subjective procedure. Developing weights in the composite equation constitutes a better approach when compared with subjective determination process. Subjectivity is probably a necessary approach to creating an index, particularly in choosing a cut-off score for performance in college, (i.e., college GPA) (Xiao, 2002). Results of the regression and numerical analyses suggest that n number of criteria can be used within a composite equation and have a specific application for the prediction of GPA.

Another important finding in this study, as other studies have shown, is that high school average is

the most important variable to predict college performance (Noble and Sawyer, 2000; Hu, 2002, Noble and Sawyer, 2002). In comparing the weight of high school average with EES or RIC, it is clear that high school average is a consistent, valid and reliable factor that is associated positively with college GPA.

CONCLUSION

The results of this study have important implications for admission officers at universities and colleges. In a broader context, it provides a method to weight criteria for admissions using the multiple regressions and the algorithmic decision-tree approach. We invite researchers to work out the algorithm in their own institutions. We would encourage others to use additional variable criteria for creating a composite equation and a number of criteria as predictors for success in college (Rowe *et al.*, 1985). Some of these factors could illuminate and can be used in composite equations. Future research might include better outcome measures, like success in work, satisfaction in college, to improve the predictive power of the admission index.

REFERENCES

Beecher, M. and L. Fischer, 1999. High school courses and scores as predictors of college success. J. College Adm., 163: 4-9.

Breland, H., J. Maxey, R. Gernand, T. Cumming and C. Trapani, 2002. Trends in college admission 2000: A report of a national survey of undergraduate admission policies, practices and procedures. ACT inc. http://www.airweb.org/images/trendsreport.pdf

Eno, D., G.W. McLaughlin, P. Sheldon and P. Brozovsky, 1999. Predicting freshman success based on high school record and other measures. Proceeding of the 38th Annual Forum of the Association for Institutional Research, May 17-20, ERIC, Minneapolis, MN., pp: 1-29.

Fralicx, R.D. and N.S. Raju, 1982. A comparison of five methods for combining multiple criteria into a single composite. Educ. Psychol. Measure., 42: 823-827. DOI: 10.1177/001316448204200315

Hu, N.B., 2002. Measuring the weight of high school GPA and sat scores with second term GPA to determine admission/financial aid index-a case study. Proceeding of the 42nd Annual Association of Institutional Research Forum, ERIC, June 2-5, Toronto, Canada, pp: 1-14.

McCormick, E.J. and D.R. Ilgen, 1980. Industrial Psychology. 7th Edn., Prentice-Hall, USA., ISBN: 013463117X, pp: 464.

Mouw, J. and R. Khanna, 1993. Prediction of academic success: A review of the literature and some recommendations. College Stud. J., 27: 328-36.

Noble, J. and R. Sawyer, 2000. Predicting different levels of academic success in college using high school GPA and ACT composite score. ACT. http://act.org/research/researchers/reports/pdf/ACT_RR2002-4.pdf

Noble, J. and R. Sawyer, 2002. Accuracy of high school grades and college admissions test scores for predicting different levels of academic achievement in college. Proceeding of the Association for Institutional Research 42 Annual Forum, June 2-5, ERIC, Toronto, Ontario Canada, pp: 1-24.

Paolillo, J.G.P., 1982. The predictive validity of selected admissions variables relative to grade point average earned in a master of business administration. Educ. Psychol. Measure., 42: 1163-1167. DOI: 10.1177/001316448204200423

Price, F.W. and S.H. Kim, 1976. The association of college performance with high school grades and college entrance test scores. Educ. Psychol. Measure, 36: 965-970. DOI: 10.1177/001316447603600422

Rowe, F.A., H. Higley, W. Larsen and D. Bills, 1985. Developing a predictive equation for admission at Brigham young university based on high school preparation. College Univ., 60: 132-141.

Sisson, H.E. and H.F. Dizney, 1980. The predictive validity of the pharmacy college admission test as compared with that of other academic predictors. Educ. Psychol. Measure., 40: 425-429. DOI: 10.1177/001316448004000221

Talley, N.R., 1989. Weighted averages and college admission. J. College Adm., 125: 19-21.

Talley, N. and J.I. Mohr, 1991. Weighted averages, computer screening and college admission in public colleges and universities. J. College Adm., 132: 9-11.

Weizman, R.A., 1982. The prediction of college achievement by the scholastic aptitude test and high school record. J. Educ. Measure., 19: 179-191. DOI: 10.1111/j.1745-3984.1982.tb00126.x

Xiao, B., 2002. Using logistic regression to find appropriate weights for a simplified academic admission index. Proceeding of the Association for Institutional Research 42th Annual Forum, June 2-5, ERIC, Toronto, Ontario Canada, pp: 1-22.

Youngblood, S.A. and B.J. Martin, 1982. Ability testing and graduate admissions: Decision process modeling and validation. Educ. Psychol. Measure., 42: 1153-1162. DOI: 10.1177/001316448204200422