Predicting Students’ Performance Through Techniques in Study Skills: A Multivariate Discriminant Analysis Approach

Shukri Shamsuddin
Faculty of Computer and Mathematical Sciences, Universiti Teknologi MARA (UiTM), Malaysia
Email: shuk927@johor.uitm.edu.my

Noor Izah Ismail
Wan Haslina Wan Hussin
Faculty of Business Management, Universiti Teknologi MARA (UiTM), Malaysia

ABSTRACT

This paper investigates some of the study skills techniques practised by students in UiTM Johor campus. The multivariate discriminant analysis approach was used to analyse the data. The discussion focuses on to what extent these study skills techniques could be used to predict the academic performance of students. This paper also examines whether there is a significant difference in study skills techniques practised by male and female students, by students of different secondary school backgrounds, and by students of different academic performances. It is hoped that this paper would shed some light on the effectiveness of the study skills program introduced to part one students upon entry into UiTM Johor.

Introduction

Students in universities and colleges have varieties of ways to study. Some students are well disciplined and follow their schedule strictly and some do not; some students might not even have any specified schedule to follow. Regardless of their styles of studying, the objective is to score high marks in their examinations.
Students are given tasks and assignments which have to be completed within a limited time. At the same time, they need to participate in co-curricular activities. It is therefore important for students to strike a balance so that they can do well in examinations as well as enjoy campus life. One might wonder how students manage their time in campus to achieve this. It is also interesting to know why some students perform well academically while others perform poorly.

There is probably no one answer to the above situation, but the differences in academic performances of students might be due to several factors such as environment, family background, formal education from rural or urban areas, exposure, study habits, learning styles, and etc.

This research focuses on identifying some of the variables that could account for the differences in students’ performances.

Problem Statement

To what extent do the variables that are described as “good” techniques in study skills discriminate between the status of students (poor, average, and excellent).

Objective

The objective of this research is to explore whether:

1. there is any difference in the range of study skills techniques employed by different gender,
2. there is any difference in the range of study skills techniques employed by "the poor", "the average" and "the excellent" students,
3. there is any difference in the range of study skills techniques employed by students from different school backgrounds,
4. a statistical model can be constructed to group the students into “poor”, “average” and “excellent” by using the multivariate discriminant analysis approach (MDA).

Theoretical Framework

Good techniques can be applied in order to improve one’s capability to do well in studies. High performers may have used different techniques of study skills as compared to the low performers. Therefore, theoretically it could be hypothesized that the academic performance of students can be discriminated by these techniques. The statistical model can be written as
\[ D = \alpha + \beta_1 X_1 + \beta_2 X_2 + \beta_3 X_3 + \beta_4 X_4 + \ldots + \beta_9 X_9 \] (1)

where the X's are the (independent) variables describing the study skills techniques and D denotes the difference in academic performance.

The equation described in (1) is known as the linear canonical discriminant function. The idea is to estimate the coefficients (b's) in the equation from a set of data to distinguish students from poor, average or excellent in academic performance, so that the groups differ in their D values (known as the discriminant scores).

In theory, probably not all of the techniques differ for students of different levels of performance. Therefore it is possible to have the best variables that discriminate the students' performance. Hence in the process of estimating b's, some of the variables might be dropped from the original equation.

For three groups, two canonical discriminant functions will be calculated. The first function will have the largest ratio of between-groups to within-groups sums of squares, and the second function is uncorrelated with the first function and has the next largest ratio (SPSS Advanced Statistics Student Guide page 96). To determine how well the functions discriminate the students' performance, the discriminant score is calculated. By using these scores, a classification table is constructed so that the probability of correct classification can be estimated.

**Hypotheses**

Apart from identifying the variables that best discriminate students' academic performance, three other hypotheses were developed as well. They are as follows:

1. There exists some differences in the range of study skills techniques employed by different genders.
2. There exists some differences in the range of study skills techniques employed by students from different school backgrounds.
3. There exists some differences in the range of study skills techniques employed by "the poor", "the average" and "the excellent" students.

**Methodology**

**Population and Sample**

The population in this study were students of UiTM Segamat Campus from semester two and above for the Nov 2001-May 2002 semester. The population size was estimated as 3,000. This included students who stayed off campus. The population was divided into three groups; students with GPAs of 2.30 and
below were categorized as academically poor, students with GPAs of between 2.31 to 2.99 were categorized as academically average and those with GPAs of 3.00 and above were categorized as academically excellent.

The sampling frame chosen was the students’ academic semester results that is, the grade point average (GPA). This sampling frame was provided by the university’s academic office. The sample consisted of 300 full-time students of which 100 were randomly selected from each group, the academically “poor”, “average” and “excellent”. A disproportionate stratified sampling was used since the proportion of students in each group was not taken into account. This is so because statistical modeling in MDA does not require proportionate sampling. Moreover, the parameter means is used in most of the analysis and the results will be more or less similar if proportionate sampling is used. Of the 300 respondents, 66 (22%) were male students and 234 (78%) were female students, which was representative of the male-female ratio of the university enrolment.

Variables and Measurements

Nine main variables were identified as good techniques in study skills from the work of Ron Fry (1991) and Gwen Gawith (1991). These variables include:

1. planning and listing according to priority,
2. keeping records on daily activities,
3. doing difficult tasks at the best time,
4. jotting down questions after lecture hours,
5. speed reading,
6. not allowing any conversations that could distract concentration while studying,
7. not allowing anything to interrupt while studying,
8. going to class early, and
9. having all important materials together while studying.

The numerical rating scale was used to measure the variables (Sekaran, 2000). Percentages were used to replace numbers in the scale as this is much more easily interpreted by the respondents. 0% was used to denote “not at all” and 100% to denote “all the time”. Respondents were required to denote the frequency, between 0 % and 100%, in which they applied the techniques described by the variables. The values of the variables were then averaged out across the board to indicate the degree of good techniques in study skills. An average of below 50% would indicate that the student is applying “poor” techniques whereas an average of 50% and above would suggest that the student is applying the “right” techniques.

The interitem consistency reliability of the nine variables was obtained. The Cronbach’s alpha coefficient was found to be 0.716. This indicates that the
internal consistency reliability of the measures used (for study skills) in this paper can be considered to be acceptable. (In general, reliabilities less than 0.6 are considered poor (Sekaran, page 312)).

Other variables that are of similar importance are gender, secondary school background, and the GPA. The latter is a general indicator of the academic performance of students. A GPA of 2.30 and below would indicate “poor academic performance”, a GPA of between 2.31 and 2.99 would indicate “average academic performance” and a GPA of 3.00 and above would indicate “excellent academic performance”. In this study the GPA is used instead of the CGPA because the variables that are being measured are only applicable for that particular semester.

Data Collection

Questionnaires were given to the randomly selected respondents. The questionnaires were given approximately two weeks after the commencement of classes. This is so because students are seen to “settle down” and “normal student life” starts to show. Two steps were taken to distribute the questionnaires to the selected respondents. The first step was to list names of the selected respondents on the students’ notice board, and a call for a meeting with the students was announced. About 40% of the students turned up for the meeting, and questionnaires were distributed and collected. The next step was to hire students to look for the remaining respondents. Overall, it took about two months to find the other 60% of the respondents to complete the data needed.

Data Analysis

The arithmetic mean and standard deviation for the nine variables was obtained. For the first hypothesis postulating differences in the range of study skills techniques employed by female and male students, a two-independent-samples t-test was performed. For the second hypothesis postulating differences in the range of study skills techniques employed by students of different school backgrounds, an F-test ANOVA was performed. For the third hypothesis postulating differences in the range of study skills techniques employed by different academic performance, an F-test ANOVA was also performed.

Finally, a multivariate discriminant analysis was performed on the students’ academic performance (dependent variable) with all of the nine independent variables.

Results and Findings

Preliminary analyses indicated that on an average, a typical student would spend 89% of the time throughout the semester having all of the important
materials while studying, 70% of the time throughout the semester going to
class early, 60% of the time throughout the semester doing difficult tasks at the
best time and 56% of the time throughout the semester practicing speed reading.
On the other hand, on an average a typical student would spend 47% of the
time throughout the semester planning and listing according to priority, 45% of
the time throughout the semester not allowing any conversations that could
distract concentration while studying, 44% of the time throughout the semester
not allowing anything to interrupt while studying, 33% of the time throughout
the semester keeping records on daily activities, and 31% of the time throughout
the semester jotting down questions after lecture hours. Upon addition of all of
the nine variables, which would indicate the overall range of study skills
techniques employed by the students, the sample indicated that on an average,
a typical student is mediocre at practising good study skills (51.2% to 54.2%
with 95% confidence level).

The sample results also indicated that there were significant gender
differences in the range of study skills techniques of the university students
(t-test; p-value of 0.001). Females in general, showed a higher application of
good study skills as compared to the male students. To check whether academic
status was associated with gender, a Chi-Square test for independence was
performed. It was found that $b^2 = 12.704$ with 2 degrees of freedom and with a p-
value of 0.002. Hence the null hypothesis that academic status is independent
of gender is rejected. The sample seems to indicate that the proportion of
female students tends to be higher than the proportion of male students in both
the average and the excellent group. Hence, this is consistent with the
application of good study skills employed by the majority of female students.

The analysis of variance also indicated that there was a significant difference
in the range of study skills techniques employed by the academically poor, the
average and the excellent students (ANOVA; p-value of 0.016). The range of
study skills techniques appeared to be significantly higher for the excellent
students compared to the poor students (t-test; p-value of 0.002), and
significantly higher for the academically average students compared to the
academically poor students (t-test; p-value of 0.044) at 5% significance level.
However the sample showed that there was no difference in the range of study
skills techniques of the average students and the excellent students.

The mean of range of study skills of students from different school
backgrounds was compared. The results of the ANOVA (p-value of 0.658) seem
to indicate that there was no difference in the range of study skills techniques
of students based upon their previous school background. The conclusion
obtained seems to contradict what some may believe that students from different
school backgrounds would have shown a different range of study skills. There
is no exact explanation of why this is so. Our guess is that students in general,
probably have somehow adapted to the "university system" after one semester.
Multivariate Discriminant Analysis

This is a two-part analysis. The first part of the analysis concentrates on all of the nine variables to predict students’ performance; whereas the second part of the analysis focuses on stepwise “wilk’s lamda selection criteria” in selecting the independent discriminatory variables.

Since there is a significant gender difference in the range of study skills techniques, the analysis in predicting student performance by discriminant functions is separated accordingly (overall, male and female). There on, comparison in correctly classified cases can be made, and the “best” discriminant functions can be chosen.

Summarization of Results for Multivariate Discriminant Analysis

There are certain assumptions made on the data for the linear discriminant functions to be “optimal”. These are;

i. the variables are jointly multivariate normal (multivariate normality), and
ii. the covariance matrices of the independent variables across the groups of dependent variable are equal (homoscedasticity).

In the analysis, only homoscedasticity was tested (Box’s M test). The null hypothesis that the covariance matrices of independent variables across the groups of dependent variable are equal is rejected for a small p-value of the Box’s M statistics. Table 1 shows the p-value of the Box’s M statistics when all of the nine variables were used as the discriminatory variables in predicting students’ performance. Table 1 also summarizes the percentages of correctly classified groups for (i) mixed sexes, (ii) only male students and (iii) only female students across the academic status group. Table 2 is similar to table 1 except that it shows the result of the stepwise method that is used to determine the discriminant functions.

From Table 1 and Table 2, the column of p-value of the Box’s M statistics indicate that Multivariate Discriminant Analysis assumptions were violated for mixed sexes and male students group when all the nine variables were used as the discriminatory variables. Hence the validity of the results is questionable. Table 1 and Table 2 suggest that prediction of students’ performance is best done by separating the gender accordingly. For the male students it is best to use the stepwise discriminant function, whereas for the female students all of the nine variables can be used.

To conclude the findings, it appears that the canonical discriminant functions can be used to predict a student’s performance. For instance, when we want to predict whether a new student would belong to any of the three groups (i.e. academically poor, average or excellent). At the end of the semester (before the final examinations), the student would be given the same
Table 1: Summarization of Correctly Classified when all the Nine Variables is used in the Canonical Discriminant Function

<table>
<thead>
<tr>
<th></th>
<th>Academically Poor</th>
<th>Academically Average</th>
<th>Academically Excellent</th>
<th>Overall</th>
<th>p-value for Box's M statistics</th>
</tr>
</thead>
<tbody>
<tr>
<td>Mixed Sexes</td>
<td>61%</td>
<td>36%</td>
<td>58%</td>
<td>51.7%</td>
<td>0.0001</td>
</tr>
<tr>
<td>Male Students</td>
<td>67.6%</td>
<td>46.7%</td>
<td>58.8%</td>
<td>60.6%</td>
<td>0.0001</td>
</tr>
<tr>
<td>Female Students</td>
<td>53%</td>
<td>27.1%</td>
<td>54.2%</td>
<td>44%</td>
<td>0.115</td>
</tr>
</tbody>
</table>

Table 2: Summarization of Correctly Classified when the Stepwise Method is used to Determine the Canonical Discriminant Functions

<table>
<thead>
<tr>
<th></th>
<th>Academically Poor</th>
<th>Academically Average</th>
<th>Academically Excellent</th>
<th>Overall</th>
<th>p-value for Box's M statistics</th>
</tr>
</thead>
<tbody>
<tr>
<td>Mixed Sexes</td>
<td>55%</td>
<td>28%</td>
<td>56%</td>
<td>46.3%</td>
<td>0.082</td>
</tr>
<tr>
<td>Male Students</td>
<td>79.4%</td>
<td>20%</td>
<td>58.8%</td>
<td>60.6%</td>
<td>0.488</td>
</tr>
<tr>
<td>Female Students</td>
<td>56.1%</td>
<td>20%</td>
<td>56.6%</td>
<td>43.2%</td>
<td>0.774</td>
</tr>
</tbody>
</table>

questionnaire (used in this research) to answer. If it is a female student, substitute all of the nine values into the canonical discriminant functions to produce the discriminant scores; the likelihood of the student belonging to a particular group can be determined. Since the overall group cases for female students that are correctly classified is 44% for the nine variables, we would expect that the chances of predicting the performance of the female student correctly is about 4 out of 10. On the other hand, in predicting the male students' performance, only one variable can be used. In substituting the value into the discriminant function to produce the discriminant score and hence classifying the students in one of the three academic status, the prediction is expected to be 61% correct.

Conclusions and Recommendations

There is no difference in the mean range of study skills techniques practised by students from different school backgrounds. Although some students (especially from urban areas) may have undergone some courses in study skills, it is of little surprise that the average students from urban areas employed the same techniques as students from rural areas while at university. In UiTM Johor, diploma students were given tips on study skills during their orientation week. Students may have employed some of the study skills techniques that
were introduced to them during their first semester in UiTM Johor and may consistently do so throughout their three years on campus.

There exists a difference in the mean range of study skills techniques employed by female and male students. A closer look at the variables revealed that variable “going to class early” and variable “having all important materials together while studying” are the techniques that contributed to the differences. For both variables, female students showed a higher mean.

There exists a significant difference in the range of study skills techniques employed by students of different academic performance. On an average, excellent students seem to spend a higher percentage of time applying some good study skills techniques as compared to the academically poor students. The results show that the excellent students tend to attempt difficult tasks at a better time more often than the poor students. They also tend to come early to class and have all the important materials around while studying more often than the poor students.

The canonical discriminant functions can be used to predict the outcome of a student’s performance. All possible discriminant functions should be produced and comparison must be made to determine the “optimal” functions that best predict the outcome. Discriminant analysis on all variables included and stepwise method of determining the discriminatory variables were employed. Assumptions on the validity of the discriminant functions must also be tested. Upon assuming that the normality of data holds true, the homoscedasticity of the data set is checked. If the data set fails the homoscedasticity test, the discriminant functions are automatically dropped. For those that passed the homoscedasticity test, percentages of original group cases that are correctly classified for each of the discriminant functions formed are then compared. The discriminant functions that produce the highest percentages of original group cases correctly classified are chosen as the “optimal” functions that best predict the outcome. In this research, the variables that were chosen to represent study skills techniques were not enough. Perhaps a similar study with more variables could be done and the “power” to predict could be improved.

References


