CHAPTER 4

FACTORS AFFECTING THE CHOICE OF COURSE OF STUDY

4.1 INTRODUCTION

In this chapter, variations in the choice of courses across gender, socio-economic characteristics and academic background will be examined. The data will be cross-tabulated to examine the socio-demographic differentials in the choice of programmes. This will be followed by multivariate analyses to examine factors affecting such choices. Appropriate statistical tests will be carried out to determine whether significant relationships exist in the bivariate and multivariate context.

4.2 CORRELATES OF THE CHOICE OF COURSE

Variation in the choice of course can be observed between male and female students. Table 4.1 shows that 60.7% of the female students and 62.8% of the male students were enrolled in the CAL programme that are based purely on external examination. Females are relatively more likely than the males to enroll in the SAM programme (33.9% versus 28.5%), but the reverse is true for "Others" programme (5.4% versus 8.7%).

Table 4.1 : Percentage distribution of students by programme according to gender, age group and parent's education level

	Programme				
	SAM	CAL	"Others"	Total	n
Gender*					
Male	28.5	62.8	8.7	100.0	172
Female	33.9	60.7	5.4	100.0	224
Age Group*					
16 to 17 years	22.0	77.1	0.8	100.0	118
18 to 19 years	36.5	55.9	7.6	100.0	263
20 years & above	21.4	35.7	42.9	100.0	14
Father's education level*					
Primary	30.4	47.8	21.7	100.0	23
Secondary	35.9	53.4	10.7	100.0	131
Tertiary	29.2	67.5	3.3	100.0	240
Mother's education level**					
Primary	33.3	50.0	16.7	100.0	30
Secondary	36.9	53.5	9.6	100.0	187
Tertiary	25.4	72.3	2.3	100.0	177

Choice of programme varies widely by age of students. Younger students are more likely than older students to enroll in CAL programmes, but the reverse is true for "Others" programme. SAM programme is relatively more popular among those aged 18-19 (See Table 4.1).

In this survey, the type of courses taken by students is found to be significantly different by parent's education level. Data show that 67.5% of students whose fathers had tertiary education were enrolled in the CAL programme. On the other hand, students whose fathers had primary or secondary education were more likely to be enrolled either in the SAM programme or "Others" programme (see Table 4.1). Students whose mothers had secondary and tertiary education tended to be more likely to be enrolled in the CAL programme. Among students whose mothers had tertiary education, 72.3% were enrolled with the CAL Programme, as compared to only 53.5% among those whose mothers had secondary education.

The choice of programme of study varies widely according to previous academic achievement and between those from Arts and Science streams. Table 4.2 shows that 21% of students from the Arts stream had enrolled in "Others" programme which is based on internally assessed courses, as compared to 2.9% among those from Science stream. On the other hand, Science stream students were more likely than those from the Arts stream to be enrolled in the CAL or SAM programmes. Since the college is well known of preparing students to further their studies in either Medicine or Engineering, it has attracted many Science students to these two programmes.

Table 4.2 : Percentage distribution of students by programme according to stream and medium of instruction

	Programme				
	SAM	CAL	"Others"	Total	n
Stream*					
Arts	19.8	59.3	21.0	100.0	81
Science	34.7	62.4	2.9	100.0	314
Medium of Instruction**					
Malay	39.2	55.3	5.5	100.0	255
English	15.2	82.8	2.0	100.0	99
Chinese	24.4	48.8	26.8	100.0	41

The type of courses taken also varies significantly according to the medium of instructions in the former schools. Table 4.2 shows that 82.8% of the English medium students were enrolled in CAL programme, as compared to 48.8% among those who came from Chinese schools and 55.3% among those who came from Malay schools. On the other hand, those from Chinese schools were much more likely to be enrolled in "Others" programme which includes Canadian Pre-University (CPU) and College Foundation Programme (TUBF).

The student's current programme of study is closely associated with their previous academic achievement. Table 4.3 shows that 57.9% of the first graders in SPM examination were enrolled in the CAL programme (which is considered the toughest course at the college), whereas 20.9% of the second graders were enrolled in courses based on internal assessment. All the SPM third graders joined

the SAM programme. Table 4.3 shows that 93.5% of the students with 'O' Level qualification had joined the CAL programme. These findings show that students with good upper secondary results were relatively more likely to join the CAL programme as compared to those with poorer results, who were relatively more likely to join "Others" programme.

Table 4.3 : Percentage distribution of students by programme according to SPM & 'O' Level qualifications

	Programme				
Qualification	SAM	CAL	"Others"	Total	n
Grade 1	38.3%	57.9%	3.8%	100.0%	261
Grade 2&3	32.8%	46.3%	20.9%	100.0%	67
'O' Level	2.2%	93.5%	4.3%	100.0%	46

Pearson Chi-Square test significant at $p < 0.05$

4.3 FITTING A MULTINOMIAL LOGISTIC MODEL FOR THE CHOICE OF COURSES

The variables in the survey are mainly categorical data. The dependent variable for this analysis is the programme in which the students were enrolled. This nominal scale variable consists of three categories, namely the SAM programme, CAL programme and "Others". Logistic regression is most frequently used to model relationship between dichotomous outcome variable and a set of independent variables. In this study, the dependent variables are polytomous in nature. According to Hosmer and Stanley (1989), in developing models for a polytomous outcome variable, one needs to be aware of its measurement scale.

Most applications involve nominal-scaled outcome variables. Let **x** denotes the vector of covariates of length p+1 with $x_0 = 1$ to account for the constant term. The polytomus logistic model is thus defined by the two logit functions, which are linear in their parameters:

$$g_1(x) = \ln\left[\frac{P(Y=1 \mid x)}{P(Y=0 \mid x)}\right]$$
$$g_1(x) = \beta_{10} + \beta_{11}x_1 + \beta_{12}x_2 + \ldots + \beta_{1p}x_p \qquad (4.1)$$

and

$$g_2(x) = \ln\left[\frac{P(Y=2 \mid x)}{P(Y=0 \mid x)}\right]$$
$$g_2(x) = \beta_{20} + \beta_{21}x_1 + \beta_{22}x_2 + \ldots + \beta_{2p}x_p \qquad (4.2)$$

where Y denotes the outcome variable, programme enrolled by students (SAM coded 1, CAL coded 2 and "Others" coded 0), and β_{ij} represents the coefficient of the independent variable x_j, factors influencing the choice of programme, in the logit functions. The above logit function $g_1(x)$ is defined by the logit transformation of the ratio of the conditional probability of **Y = 1** (student joining SAM programme) given the independent variables x_j to the conditional probability of **Y = 0** (student joining "Others" programme) given the same independent variables. Whereas the logit function $g_2(x)$ is defined by the logit transformation of the ratio of the conditional probability of **Y = 2** (student joining CAL programme) given the independent variables x_j to the conditional probability of **Y = 0** (student joining "Others" programme) given the same independent variables.

It follows that the three conditional probabilities of each outcome category given the covariate vector are:

$$P(Y=0 \mid x) = \frac{1}{1+e^{g_1(x)}+e^{g_2(x)}} \tag{4.3}$$

$$P(Y=1 \mid x) = \frac{e^{g_1(x)}}{1+e^{g_1(x)}+e^{g_2(x)}} \tag{4.4}$$

and

$$P(Y=2 \mid x) = \frac{e^{g_2(x)}}{1+e^{g_1(x)}+e^{g_2(x)}} \tag{4.5}$$

Further, the log-likelihood function is given by

$$L(\beta) = \sum_{i=1}^{p} y_{1i} g_1(x_i) + y_{2i} g_2(x_i) - \ln(1+e^{g_1(x_i)}+e^{g_2(x_i)}) \tag{4.6}$$

where $\sum y_{ij} = 1$ for each i.

The likelihood equations are found by taking the first partial derivatives of $L(\beta)$ with respect to each of the 2(p+1) unknown parameters β_{jk} where j = 1,2 and k = 0,1,2,...,p. The maximum likelihood estimators β_{jk} 's are obtained by setting these partial derivatives to zero and solving for β_{jk} 's .

4.3.1 MODEL-BUILDING : VARIABLE SELECTION

The criteria for inclusion of a variable in a model may vary from one problem to another. The traditional approach to statistical model building involves seeking the most parsimonious model. The univariate was presented in Table 4.4. The analysis was performed by using the statistical software package SPIDA (Statistical Package for Interactive Data Analysis). Any variable with p-value (Wald-Test) less than 0.25 is considered as candidate for multivariate model along with all variables of known importance. Use of the 0.25 level as a screening criterion for selection of candidate variables is based on the work by Bendel and Afifi (1977) on linear regression and on the work by Mickey and Greenland (1989) on logistic regression. These authors showed that use of a traditional level (0.05) often fails to identify variables known to be important.

The variables chosen as candidates for multivariate regression are Gender, Age group, Mother's education level, Medium of instruction in the previous school, Stream (Arts or Science), Region of residence. The variables "Age group" and "Region of residence" were included even though some levels were tested to have p-value more than 0.25. But these variables are retained in the model because they are considered important factors in choosing the programme of study. Contrary to expectation, the variable "SPM result" and " Father's education level" were found to be insignificant. It is to be mentioned that the college under study enrolled students based on their forecast results supplied by the secondary schools attended by the students. Furthermore, the forecast results might differ from the final results.

Table 4.4 : Univariate test (Wald Test) with reference to logit function $g_1(x)$ in (4.1) & $g_2(x)$ in (4.2) on choice of courses

Variable	Logit $g_1(x)$ Coefficient β_{jk}	Std Error	p-value	Logit $g_2(x)$ Coefficient β_{jk}	Std Error	p-value
Gender = 1 (Male)#						
Gender = 2 (Female)	0.775	0.614	0.206	0.562	0.600	0.349
Mother Education Level = 1 (Primary)#						
Mother Education Level = 2 (Secondary)	1.253	0.699	0.073	1.263	0.676	0.062
Mother Education Level = 3 (Tertiary)	9.824	31.894	0.758	10.533	31.893	0.741
Medium = 1 (Malay)#						
Medium = 2 (English)	-0.798	0.862	0.354	0.532	0.828	0.521
Medium = 3 (Chinese)	-1.897	0.726	0.009	-1.547	0.688	0.025
Stream = 1 (Arts)#						
Stream = 2 (Science)	1.919	0.636	0.003	1.444	0.600	0.016
Region = 1 (Southern)	0.433	1.095	0.692	0.139	1.082	0.898
Region = 2 (Central)#						
Region = 3 (Northern)	-0.548	0.849	0.519	-1.288	0.845	0.127
Region = 4 (East Malaysia)	0.226	1.102	0.838	0.011	1.085	0.992
Region = 5 (Foreign country)	4.154	25.732	0.872	5.493	25.715	0.831
Age group =1 (16 – 17 years)	7.405	27.621	0.789	8.245	22.620	0.765
Age group =2 (18-19 years)#						
Age group = 3 (> 19 years)	-2.773	0.848	0.001	-2.674	0.762	0.000

reference categories.

4.3.2 TESTING FOR SIGNIFICANCE OF MODEL AND GOODNESS-OF-FIT

As the independent variables in the logistic model are inter-related, the importance of each variable included in the model would be assessed net of the effects of other variables. The statistic used in assessing the significance of an independent variable in the polytomous logistic model is thus the likelihood ratio G which is defined by

$$G = -2\ln\left[\frac{\text{likelihood without the variable}}{\text{likelihood with the variable}}\right] \tag{4.7}$$

Under the null hypothesis, the coefficients are zero, and the G statistic will follow a Chi-Square distribution. In general, the likelihood ratio test for significance of coefficients for a variable will have degree of freedom equal to the number of outcome categories minus one times the degrees of freedom for the variable in each logit. Table 4.5 shows the G statistic when the model with one variable is compared to model with constant term only. Table 4.6 illustrates the results of fitting the polytomous logistic regression with selected variables from section 4.3.1, entered one at a time. The G statistic and the corresponding p-value are given to show the significance of the variable.

Table 4.5: Results of fitting one variable into the constant model

Variables	-2log L	df	G statistic	df	p-value
Constant	576.392	376			
Gender	574.323	374	2.069	2	0.355
Medium of Instruction	548.729	372	27.663	4	0.000
Stream	567.384	374	9.008	2	0.011

The variables "Mother's Education Level", "Region of residence" and "Age group" were tested insignificant ($p > 0.05$), and dropped from subsequent analyses. But the variable "Gender" was included as this was found to be an important factor in the previous studies. Hence, the final model consisted of three independent variables namely, Gender, Stream and Medium of instruction in the previous school. The results of the fitting the data to the polytomous logistics regression model are presented in Table 4.6. Both the odds ratio and confidence interval are included for further interpretation of the model.

Table 4.6: Results of fitting polytomous logistic regression to the survey data with reference to logit function $g_1(x)$ & $g_2(x)$ on choice of courses

Logit	Variable(x_{ij})#	Coeffi-cient β_{ik}	Std. Error	p-value	Odds	Confidence interval
$g_1(x)$	Constant	0.814	0.697	0.243		
	Gender = 1 (Male)#					
	Gender = 2 (Female)	0.911	0.644	0.157	2.488	(0.704 , 8.792)
	Medium = 1 (Malay)#					
	Medium =2 (English)	-0.166	0.915	0.856	0.847	(0.147, 5.093)
	Medium =3 (Chinese)	-1.722	0.751	0.022	0.179	(0.041, 0.779)
	Stream = 1 (Arts)#					
	Stream = 2 (Science)	1.995	0.689	0.004	7.352	(1.905, 28.374)
$g_2(x)$	Constant	1.402	0.667	0.036		
	Gender = 1 (Male)#					
	Gender = 2 (Female)	0.553	0.630	0.380	1.739	(0.506, 5.975)
	Medium = 1 (Malay)#					
	Medium =2 (English)	1.207	0.884	0.172	3.343	(0.591, 18.895)
	Medium =3 (Chinese)	-1.408	0.713	0.048	0.245	(0.061, 0.990)
	Stream = 1 (Arts)#					
	Stream = 2 (Science)	1.944	0.662	0.003	6.990	(1.910, 25.579)

#Reference categories

The – 2 log likelihood L with the variables included in the polytomous logistic model was found to be 537.228 (computed by the SPIDA software) with 368 degrees of freedom. The –2log likelihood L's value for model without any covariate was 576.392 with 376 degrees of freedom. Hosmer and Stanley (1989) suggested the approximate R^2 type measure to be used in assessing the goodness-of-fit of a logistic regression model. Let L_o & L_p denote the log-likelihood for models containing only the constant, and the model containing the constant plus the p covariate, respectively. The form of statistic proposed by them for use in logistic regression was given by:

$$R_L^2 = \frac{100(L_0 - L_p)}{L_0} = 100(1 - \frac{L_p}{L_0}) \qquad (4.7)$$

From the above, the log-likelihood values for the suggested polytomous logistic model, the value of the statistic was calculated to be $R_L^2 = 6.8$. This figure shows that the model explains approximately 6.8% of the variations in the data of student's choice of programme. The polytomous logistic regression model obtained in this study is defined by the two logit functions:

$$g_1(x) = 0.814 + 0.911GE(2) - 0.166MD(2) - 1.722MD(3) + 1.995SM(2)$$

and

$$g_2(x) = 1.402 + 0.553GE(2) + 1.207MD(2) - 1.408MD(3) + 1.944SM(2)$$

where the abbreviation for the variables are defined in Table I in the Appendix III.

The positive coefficients indicate an increase in the log odds and consequently the increase in the probabilities of students with the given characteristics joining SAM or CAL rather than the "Others" programme when compared to the reference categories. The negative coefficients indicate a decrease in the probabilities of students with the given feature joining SAM or CAL rather than the "Others" programme when compared to the reference categories.

4.3.3 ANALYSIS OF FACTORS AFFECTING CHOICE OF COURSES

Table 4.6 summarizes the results of the polytomous logistic regression model described above. Several variables in the polytomous logistic model are of great interest to the management of the College. The estimated odds ratios and 95% confidence interval for the variables gender, medium of instruction and stream are given in Table 4.6. In the following section, the interpretations of various coefficients of the variables will be discussed further.

(i) GENDER

In Table 4.6, the positive coefficients of Gender = female indicate an increase in log odds and consequently an increase in the probability of female students joining SAM or CAL as against "Others" programme when compared to male students. Controlling for other variables, the odds of female students joining the SAM Programme rather than "Others" programme is 2.488 that of the male students. This figure shows that female students are more than twice as likely as the male students to join SAM rather than the "Others" programmes. However,

with reference to the second logit function, female students are only 1.739 times more likely than their male counterparts to join CAL rather than "Others" programmes.

(ii) MEDIUM OF INSTRUCTION IN SECONDARY SCHOOL

Table 4.6 shows students from English and Chinese Medium schools have odds ratios of 0.847 and 0.179 respectively as compared to those from Malay medium schools. These figures imply that at 95% confidence level, students from Malay medium schools are 1.2 and 5.5 times more likely than those from English and Chinese medium respectively to join SAM rather than "Others" programme. Whereas students from English medium schools are 3.3 times more likely than those from Malay medium to join CAL rather than "Others" programme. But Malay medium students are 4 times more likely than those from Chinese medium students to join CAL than "Others" programme.

(iii) STREAM (Arts or Science)

Controlling for other independent variables, Science stream students are approximately 7.3 times more likely than those from the Arts stream to choose SAM rather than "Others" programme. Science stream students were also found to be 7 times more likely than Arts students to choose CAL rather than "Others" programme (See Table 4.6).

4.4 FOCUS GROUP DISCUSSION ON THE CHOICE OF COURSES

Three focus group discussions (FGD) were carried out on students in the private college. The students were randomly selected. The breakdown of students for the FGD from various programmes was shown in Table 2.5. Questions related to the choice of course and college processes were asked. The summary of discussions was presented in the next two sections. All students from the focus group discussions were asked when they first started thinking seriously about going to a private college to pursue higher education. The results indicate most of the students had started thinking about pursuing their education in a private school after completing their secondary education. Those who intended to pursue a degree in USA had started thinking about which college to go as early as after primary education. A good majority of them (38 out of 43) mentioned that their parents have thought of all possible options of their education when they were in primary schools. The breakdown of the percentage of students on when they started to think about a private college is shown in Table 4.7.

Table 4.7: Distribution of percentages for students started thinking of further studies in a private college

Level	SAM	CAL	"Others"
After Primary	2 (11.8%)	3 (15%)	4 (80%)
After Lower secondary	14 (82.3%)	15 (75%)	1 (20%)
After Upper Secondary	1 (5.9%)	2 (10%)	0 (0%)
Total	17(100%)	20(100%)	5(100%)

Some of the comments made by students are summarized below:

David, a CAL student: Why? I always look forward to a college that is excellent in its academic performance so that I would somehow be guaranteed to obtain good results to further my studies overseas. High standard and good reputation are my criteria of choice.

Michele, a SAM student: David is right. I have to make sure that I choose a well-established and recognized college. To add on to what he said, I personally feel that since I am from the Malay medium school and I really need to choose a course that suited me so I'll not face any difficulty in the course.

George and Elaine, both from English medium school commented: We don't face this problem. As almost all courses conducted in private colleges are in English medium. But we are more concerned about how relevant the course we are taking as we wish to pursue an engineering degree in UK. On top of that, we need to know whether the course is fully recognized by the universities.

Peter, from ADP programme: My parents have made the arrangement for me right when I was in primary school. They told me that American education is broader and both of them had graduated from one of the American universities. I feel that they just wanted me to do a course that leads me to an American degree.

Joshua answered: I live in Subang Jaya and certainly I choose a college nearby to avoid traffic jam.

Ong, Jason and Poh Lian, three SAM students said: Tuition fee and cost of living are the two main factors to us, as we are from outstation. Although the course is really expensive but thinking of the good quality of education provided by the college, our parents have no choice but to enroll us.

When the issue on workload and entry requirement were discussed,

Jane: Yes, I found out from my sister that SAM and CAL are too intensive, heavy workload so I chose other programme.

Lee, CAL student: I learnt from my friends that CAL is the toughest course in the college but universities worldwide recognizes it. I know it would be tough but I still took up the challenge.

Tan and Lim, both from SAM commented: The entry requirements of courses offered by private colleges are almost the same and quite easy to meet. We were not worried at all when we came in to put in our applications, knowing that we will definitely get a place

4.5 CONCLUDING REMARKS

The survey data indicate that the factors that explain the differentials in the choice of course are related to the family background or the academic background of the students. The probability of joining certain programmes according to selected characteristics can be ascertained from the regression models. However, it must be mentioned that the predictive power of the models is rather weak, as some important variables were not included in the polytomous logistic regression model. Information pertaining to parent's income and measures of quality education were considered the

sensitive issues hence were not collected in the survey. Be that as it may, findings from the bivariate, multivariate analyses as well as FGDs are rather consistent and this shows that the results are quite reliable.