Throughput of UWC students who did at least one semester of third-year Statistics

by

Abduraghiem Latief

A minithesis submitted in partial fulfillment of the requirements for the degree of Magister Scientiae in the Department of Statistics, Faculty of Science, University of the Western Cape.

Supervisor: Prof. R. Blignaut

Co-supervisor: Prof. D. Kotze

October 2005

Throughput of UWC students who did at least one semester of third-year Statistics

Keywords

Education

Throughput rate

Statistics

Completion rate

Pass rate

Undergraduate

University studies

First time entrants

Graduation

Enrolment

Abstract

Throughput of UWC students who did at least one semester of third-year Statistics

The study explores the completion rates (the number of years a student takes to complete a degree) of graduates at the University of the Western Cape (UWC) in South Africa. The graduates in the study all did at least one semester of statistics in their final year of study. The students' completion will be described with respect to school results and socio-demographics. Differences between students who finished their studies in the prescribed time of three years and those who took longer than the prescribed time will be highlighted.

Factors that aid or hinder students from successfully completing their studies in the prescribed time will be analyzed. An entry selection model will be developed to screen the students. This will assist with an enrolment strategy.

The most significant result found was that the political environment played the most significant role in throughput. The next significant result from the study showed that the grade 12 aggregate played a significant role in throughput. It is suggested that UWC be proactive in developing alternative methods of selecting students, since the

new Further Education Training (FET) school system, which will be implemented in 2006, will omit the grade 12 aggregate.

October 2005

Declaration

I declare that ***Throughput of UWC students who did at least one semester of third-year Statistics*** is my own work, that it has not been submitted for any degree or examination in any other university, and that all the sources I have used or quoted have been indicated and acknowledged by complete references.

Abduraghiem Latief October 2005

Signed: ....

List of tables

List of figures

Abbreviations

ASCII	American Standard Code for Information Interchange
CESM	Classification of Educational Subject Matter
CHE	Christian Higher Education
DoE	Department of Education
DVN	Diploma in Veterinary Nursing
EXCEL	Microsoft spreadsheet
FET	Further Education Training
HDI	Historically Disadvantaged Institutions
HE	Higher Education
HEIs	Higher Education Institutions
HWI	Historically White Institutions
ICS	Information and Communication Services
MPM	Mean percentage mark
PoE	Place-on-exam
PRN	Print file format
SAPSE	South African Post-Secondary Education
SAS	Statistical Analysis Systems
SAUVCA	South African Universities' Vice Chancellors' Association
SPSS	Statistical Package for Social Science
UWC	University of the Western Cape

Acknowledgements

I would like to thank my parents, my wife (Shaheema), my supervisors (Professor R. Blignaut and Professor D. Kotze), the Madsens from USA and colleagues for their patience and support. Lastly I thank the Almighty for giving me the strength and health to finish this thesis.

Contents

Chapter 1

Introduction

1.1 Background to the study

The Department of Education (DoE) restructured the higher education (HE) system in 2000 (Asmal, 1999). The restructuring caused universities to re-align themselves with the priorities of the Department of Education. In 2001 the Department of Education introduced its new five-year national plan for higher education. One of the priorities in the national plan was to increase undergraduate output to ensure that the current demand for high-level managerial and professional skills be met (Department of Education, 2001a).

This priority to increase undergraduate output (also known as undergraduate throughput) initiated the study described in this mini-thesis. This study explores the throughput rate of UWC students who did at least one semester of third-year level Statistics. *Throughput* is the number of undergraduates who complete their studies in the prescribed time (Cairncross, 1999). The throughput is one of the factors that the government uses for funding a university (Department of Education, 2001b).

The Department of Education has introduced a new funding formula which is applicable to all higher education institutions (HEIs). This new formula takes undergraduate output as a factor in determining the funding that a university will receive (Department of Education, 2001b). The previous South African post-secondary education (SAPSE) formula was based on four criteria:

1. Student numbers - the overall number of students.
2. Area of study - for example humanities/science.
3. Student throughput - the pass rate.
4. Level of studies - honour's level is equal to two times the undergraduate level, master's level is three times the undergraduate level and doctoral level is equal to four times the undergraduate level.

The South Africa post-secondary education (SAPSE) funding formula favoured the historically-white institutions (HWI) more than the historically-disadvantaged institutions (HDI), which had high failure rates, few science students and low postgraduate student numbers (The Mail and Guardian, 1999).

1.2 Statement of the problem

The study will investigate the throughput of students who did at least one semester of third-year level Statistics in the Department of Statistics at the University of the Western Cape (UWC). Completion of undergraduate studies by a student in three consecutive years will be defined as successful throughput. The study explores factors that could contribute to students successfully completing their studies in the prescribed amount of time.

1.3 Purpose of the study

The school system under-prepares students for higher learning (Nair, 2002). This is worsened when they enter into higher institutions. In other words, it leads to low throughput rates. O' Connell (2004) indicated that UWC's throughput rate is 17% for the whole university. This study will describe the throughput rate of a subset of students from UWC and explore some factors that might contribute to throughput.

1.4 Aim of the study

The aim of the study was to explore the throughput rate of third-year Statistics students in the Department of Statistics and to model the probability of successful throughput with certain factors or predictor variables. The following factors were explored: gender, race, home language, Grade 12 aggregate, Grade 12 mathematics results, entering university directly after school and student registration before and after the 1994 elections in South Africa (first democratic election).

1.5 Research questions of the study

The goal was to identify what factors influence successful throughput. Various modelling techniques were used to identify the factors that significantly predict successful throughput. Logistic regression and decision trees were used. The aim was, furthermore, to establish if the change in the political arena, specifically the change after the democratic elections in 1994, had an influence on successful throughput.

1.6 Importance of the study

By improving throughput, more skilled students will become quality scientists and employees. The benefits of increasing undergraduate output are:

1. More successful students will enter into the job market and promote UWC in their company profiles (as alumni).

2. More undergraduate students will be available from whom to recruit for postgraduate studies.

3. If students complete their studies in the prescribed time, they will save on tuition fees. The university will gain by earning its subsidy more quickly.

4. More funds will be forthcoming for research and postgraduate studies.

The information from this study can aid in improving enrolment strategies at UWC.

1.7 Outline of the study

The following topics will be dealt with in each chapter:

Chapter 2 is a literature review. In this chapter, the views of other researchers on throughput and the factors that contribute to it are expressed.

Chapter 3 describes the methodology: How the sample was collected, what population was of interest and what objectives were defined.

Chapter 4 presents the findings of the analysis. First, descriptive results are shown, followed by univariate and multivariate logistic regression analysis. Decision-tree results are summarized.

Chapter 5 discusses and interprets the findings. This chapter concludes with suggestions and recommendations.

The next chapter will deal with the views of some scholars on the topic of throughput.

Chapter 2

Literature review

In the previous chapter, background information as to how the study evolved was described. In this chapter, literature on how others view the problem of throughput and what factors they believe contribute towards throughput will be shown. Different studies investigated different factors within specific programmes. To make the literature comparable, emphasis is placed on those factors that are important to this study. The chapter deals with the demographic background (gender, race and home language), school background (aggregate and mathematics) and the political environment under which the students studied.

Fraser and Killen (2003) use the term *academic success* to indicate that students are able to meet the assessment requirement of the programme in which they enroll; if these requirements can be met in the minimum time, that represents greater success than if subjects have to be repeated. Bitzer and Troskie-De Bruin (2004) argue that throughput and completion rates should not be seen as the only criteria of quality or the hallmark of high standards.

Cairncross (1999) investigated the fourth/final-year level Human Ecology students' throughput rates and completion rates. She defines throughput rate as "the number of students who pass through a period in the allocated time period" (p. 2). The throughput in the case of the Human Ecology students is the number of students who completed their studies in four years. She also defines completion rate as "the number of students who complete their studies" (p. 3). She mentions that, historically, Grade 12 results are used to categorise students into those who qualify for degree courses and those who qualify for diploma courses. She refers to a student who leaves a course as a "dropout" (p. 2). She does recommend that dropout students be referred to as "early exits" (p. 2). Table 2.1 below summarises her findings in terms of the following categories – the overall findings (all the students together), the Human Ecology student registration for the years 1994 to 1996, the Human Ecology degree course and the Human Ecology diploma course.

Table 2.1 **Throughput and completion rate results of Cairncross's study**

	Overall	1994	1995	1996	Degree	Diploma
Throughput rate	37 (33.9%)	9 (25.7%)	12 (32.4%)	16 (42.2%)	38.8%	16.7%
Completion rate	18 (16.5%)	7 (20%)	11 (29.7%)	0 (0%)	17.6%	12.5%
Dropout rate	54 (49.54%)	19 (54.3%)	14 (37.8%)	21 (56.8%)	43.5%	70.8%

(Source: Cairncross, 1999)

Since the entrance requirements are different for a degree and a diploma, it can be seen that the throughput rates are higher for the degree course compared to the diploma course in Human Ecology. For the individual years of registration from 1994 to 1996, the throughput rates increased. In 1996 more than half of the students dropped out of the Human Ecology course. This means that all the students who remained in the course completed it in the prescribed amount of time. For the degree course (Grade 12 exemption) the throughput rates were more than double those of the diploma course. The dropout rate for the diploma course was higher than that observed for the degree course.

The quality and characteristics of students at different universities were investigated by Taylor and Harris (2002). They derived their data from the South African post-secondary education (SAPSE) information system. Those universities whose South African post-secondary education (SAPSE) databases were incomplete were excluded from their study; therefore, they could only include ten universities in their study. The ten universities are the University of Cape Town, the University of Durban-Westville, the University of the Orange Free State, the University of Port Elizabeth, Potchefstroom University for CHE, the University of Pretoria, Rand Afrikaans University, Rhodes University, the University of Stellenbosch, and the University of Zululand. Taylor and Harris (2002) define a university to be efficient if it complies with the following definition of efficiency: "Efficiency involves minimizing the inputs required to produce a given output or, conversely, maximizing the output from given inputs" (p. 184). The input measure for a university includes students, personnel and financial resources. The output is graduates and research production. Dawes P., Yeld N. and Smith M.J. (1999) state that the national enrolment goals will be linked to funding in future. Graduate output was a factor in the old funding system (SAPSE) and it will also be a factor in the new funding system.

2.1 Demographic background

Dawes et al. (1999) express the need to increase the participation rate of black Africans in higher education (HE). They mention that black Africans are being disadvantaged in the selection system for higher education because of unequal schooling (study under unfavourable and disadvantaged conditions). They mention that access and admission to higher education will become more difficult for black

Africans because more black Africans will enter with low aggregates. They encourage the investigation of race and gender to see if there is an increase in enrolment.

Nair (2002) defines throughput rate as "number of years used by many students to complete a degree or diploma" (p.98). Nair relates low throughput to under-preparedness due to the inadequate schooling system. Nair gives the national average of the throughput rate in HEIs in Table 2.2 as follows:

Table 2.2 **Throughput rate at HEIs nationwide**

Throughput rate (%) for population groups	
African students	8
White students	25
Throughput rate (%) in key subject areas for African students	
Engineering	3
Medicine	9
Natural Science	12

(Source: Frank Meintjies: Deloitte Consulting, taken from Nair, 2002)

Lourens and Smit (2003) built a predictive model to predict the success of students in their first-year level of studies. Lourens and Smit (2003) used the following demographic background predictors - age of student, province of matriculation, Grade 12 aggregate, Grade 12 English symbol (defined as adequate or inadequate), ethnic group, gender, campus of study (Pretoria campuses versus satellite campuses), method of study (full-time versus part-time), financial aid (yes or no), marital status, type of accommodation (resident student or not) and classification of educational subject matter (CESM), i.e. major field of study, to describe the type of students entering Technikon Pretoria. Lourens and Smit (2003) divided the English grade symbols into two groups, namely, the "adequate" group - higher grade D symbols or better, the standard grade - C symbols or better and the lower grade - B symbols or better. The rest were in the "inadequate" group. They made use of stepwise logistic regression to find the model with the most significant predictors. Lourens and Smit (2003) confirmed a relationship between school aggregate and first-year success rate. Lourens and Smit (2003) also found that a relationship does not exist for second and third-year successes. They used eight significant independent variables in the study to build two models. The first model consisted of all eight variables and the other model only consisted of the CESM category and the Grade 12 aggregate. They then compared the performance of the two models. They concluded that both models have more than a 70% predictive accuracy and that the Grade 12 aggregate and major field of study play an important role in terms of students' first-year success at Technikon Pretoria.

Van Rooyen (2001) found that English as a home language was a significant predictor of the bridging-year mean percentage mark (MPM). Agar (1991) confirmed that disadvantaged students found it difficult to express themselves in English. He found that 75.3% of students in a bridging programme at the University of the Witwatersrand attribute the difficulties of academic actualization to language barriers. Howie (2003) confirms these views, showing that pupils' English proficiency was a strong predictor of success in mathematics.

2.2 School background

Nair (2002) states that the government loses millions on students who fail at higher education institutions (HEIs) and also spends millions on a schooling system which produces school leavers who are under-prepared for higher education and the job market. Keeping this in mind, Botha A.E., McCrindle C.M.E. and Owen J.H. (2003) state:

> "In the South African education system, students write a standardized, independently set, matriculation examination at the end of their school career (Grade 12). The results of this examination are used as the main criteria for admission to tertiary educational institutions. Subjects may be taken on two levels –higher grade and standard grade. A proposed new matriculation curriculum, however, will eliminate the difference between the standard and higher grades" (p. 132).

The Diploma in Veterinary Nursing (DVN) programme uses Grade 12 mathematics with its grades as a selection criterion (Botha et al., 2003).

Table 2.3 **Scoring system used by the University of Pretoria**

Matriculation symbol	Higher grade	Standard grade
A (more than 80%)	5	4
B (70 – 79%)	4	3
C (60 – 69%)	3	2
D (50 – 59%)	2	1
E (40 – 49%)	1	0

(Source: Botha, McCrindle & Owen, 2003).

Botha et al. (2003) define the adjusted mark as "standard grade minus 10%" (see Table 2.3) and set the minimum of 40% of the adjusted mark for both higher grade and standard grade. They used the Mann-Whitney non-parametric test to test for the difference between groups (p-value = 0.0097) and found that a statistically significant difference does exist in the adjusted mark obtained for Grade 12 mathematics between the groups that passed and failed the first-year veterinary nursing course. This means Grade 12 mathematics is related to success or failure of veterinary nursing students at tertiary level. They recommend that students with Grade 12 mathematics marks higher than 57% be given preference for admission to veterinary nursing courses. Therefore mathematics can be used as an admission criterion for enrolment for a veterinary nursing course.

Table 2.4 **First-year level result versus Grade 12 mathematics grade of diploma course students in Veterinary Nursing**

Result	Higher grade	Standard grade
Pass	12	26
Fail	48	56

(Source: Botha, McCrindle & Owen, 2003)

Botha et al. (2003) found that no statistically significant relationship (Table 2.4 gives a Chi-square p-value = 0.1196) exists between the grade of mathematics at matriculation level and the success or failure in the first-year level of study.

The following people oppose the view that Grade 12 mathematics is a significant factor in successful completion of tertiary education. Mitchell (1988) says that there is no significant difference between those students who did Grade 12 mathematics and those who did not do Grade 12 mathematics, with respect to an accounting degree, excluding the quantitative courses. Bargate (1999) also found that Grade 12 mathematics did not play a significant role in overall academic performance.

Dawes et al. (1999), in their study, used the aggregate school score, which is the raw total of all the marks for all a student's school subjects. They then define a place-on-exam (PoE) indicator by taking the individual aggregate school score for all the students at a particular school and assigning the indicator to that rank score. They give three reasons for the advantage of using the place-on-exam. Firstly, scores are compared within the same school, so students will not become victims of circumstances. Secondly, it can be used as a measure of relative merit for students without it being influenced by the examination system or internal assessments at the school. Thirdly, it is easy to use and interpret (Dawes et al., 1999). Dawes et al. (1999) say that in a study done by Stoker D.J., Engelbrecht C.S., Crowther N.A.S., Du Toit S.H.C. and Herbst A. (1986), it was found that aggregate score was the strongest single predictor of success at university. Dawes et al. (1999) also state that other South African studies done by Skuy M., Zolessi S., Mentis M., Fridjhon P. and Cockcroft K. (1996); Badenhorst F.D., Foster D.H. and Lea S.J. (1990) and Shochet (1985) support Stoker et al. (1986)s' findings, but they did not focus on race or gender. In some studies (Badenhorst et al., 1990; Shochet, 1985) in which race was investigated, the sample of Blacks was too small to deduce information regarding race as predictor. Where the sample size was large, the results between school examination and success at university were too complex to understand (Dawes et al., 1999).

Lourens and Smit (2003) found, in their study, that Grade 12 aggregate and major field of study were the most important predictors for the success of students in their first-year level of study. They found, in their study, that only 20.96% (1016 out of 4848) of first-years passed all their subjects first time around.

2.3 Political environment

Taylor and Harris (2002) investigated the efficiency of the following universities: the University of Cape Town, the University of Durban-Westville, the University of the Orange Free State, the University of Port Elizabeth, Potchefstroom University for CHE, the University of Pretoria, Rand Afrikaans University, Rhodes University, the University of Stellenbosch, and the University of Zululand. Taylor and Harris (2002) found that the student numbers increased, from 1994 to 1997 for the ten universities, by an average of 4.7% (compound rate) per annum. They state that a university with high student numbers is generally associated with improved university efficiency. But the academic successes of students have no relationship to the efficiency of a university (Taylor & Harris, 2002).

In the next chapter, the research design and methodology will be discussed.

Chapter 3

Research design and methodology

3.1 Statement of the problem

Completion of undergraduate studies by a student in three consecutive years will be defined as successful throughput in this study. The study explores factors or predictor variables that could contribute to students successfully completing their studies in the prescribed amount of time.

3.2 Aim of the study

The aim of the study was to model the probability of successful throughput with certain factors. The following factors were explored: gender, race, home language, Grade 12 aggregate, Grade 12 mathematics results, entering university directly after school and student registration before and after the 1994 elections in South Africa (first democratic election).

3.3 Objectives of the study

The objective of the study was to investigate the relationship between the factors (mentioned below) and successful throughput. The factors considered were: gender, race, home language, Grade 12 aggregate, Grade 12 mathematics results and time between school and university. Afterwards various modelling techniques (logistic regression and decision trees) were used to identify the factors that significantly predict successful throughput. The aim was, furthermore, to establish if the change in the political arena, specifically the change after the democratic elections in 1994, had an influence on successful throughput.

3.4 Hypotheses

The following hypotheses were tested:

1. Females were more likely to complete their studies in the prescribed time than males.
2. African students were less likely to complete their studies in the prescribed time than non-African students.
3. Students who speak English as a home language were more likely to complete their studies in the prescribed time than non-English home language speaking students.
4. The throughput rate of students with Grade 12 aggregate symbols less than 60% was lower than those with Grade 12 aggregate symbols of 60% and above.
5. The throughput rate of students with Grade 12 mathematics symbols less than 60% was lower than those with Grade 12 mathematics symbols of 60% and above.
6. A relationship exists between throughput and a break between school and university studies.

3.5 Study design

The study design was a historical cohort (retrospective) study because historical student records were used. The cohorts under consideration were those who completed their studies within three years versus those who took longer than three years. The events like registration and completion of academic studies occurred prior to the start of the study.

3.6 Study population

The population for this study consisted of all students who had completed at least one semester of either Mathematical Statistics or Applied Statistics at the third-year level in the Department of Statistics at the University of the Western Cape. It did not matter if the student had failed the semester or repeated the semester in the next academic year. Students who registered for both semesters and obtained zero for both semesters were omitted from the study. The students who obtained zero for both semesters either did not deregister for the course or stopped attending lectures and completed no assignments, tests or exams. All transfer students from other institutions where omitted from the study. Transfer students, are students who have finished some of their subjects or academic year levels at an institution other than UWC, and then come and registered at UWC to continue their studies.

The entrance requirements for students to study Statistics at UWC are:

1. A matriculation exemption certificate issued by the Matriculation Board of the South African Universities' Vice Chancellors' Association (SAUVCA);
2. A pass of at least 40% in the higher grade or 50% in the standard grade examination for Mathematics; and
3. A pass of at least 40% in the higher grade or 50% in the standard grade examination for either Biology or Physical Science; or
4. An examination recognized by the Joint Matriculation Board for this purpose.

The duration of a B.Sc. degree in the Science Faculty at UWC is three years, with a time limit of five years for full-time study. Furthermore, the student has to obtain a minimum of 360 credits to obtain the degree. To major in Statistics, a student either starts in his/her first-year level with Statistics and then follows it through to third-year level, or starts in the second-year level and continues to third-year level. The option of starting from second-year level depends on the student passing first-year level university Mathematics (Science Faculty, 2004).

In 1987 and 1990, UWC introduced Mathematical Statistics and Applied Statistics up to a third-year level, respectively. Both courses were divided into two semesters. Students should pass both semesters of Mathematical Statistics or Applied Statistics at third-year level to major in Statistics. The Applied Statistics course contained theoretical elements with its application, but less emphasis is placed on mathematical ability. In 2002, the two courses were combined for various reasons, none of which are relevant for the purpose of this study. Data were collected for these two courses from 1975 to 2004.

3.7 Sample size

Data on all students who completed at least one semester of third-year Statistics were collected. In total, 409 students met the criteria for inclusion.

3.8 Data collection

Data for this study were historic (retrospective) and were collected from the university records. The data were extracted internally from the UWC's mainframe database (secondary source) without the need for a research instrument. The database is maintained by UWC's Information and Communication Services (ICS). All student data generated during the normal academic enrolment, such as registration, student marks, year of completion, year of graduation, et cetera, were captured. The data were then stored in an ORACLE mainframe student database. With the permission of ICS, any academic staff member can request information regarding his or her students for research purposes.

Requested data can either be in paper or electronic format. The data were electronically mailed as an attachment in a print file format (PRN). The data were then imported into EXCEL. The variable names were assigned in EXCEL; for example, the variable name *studnum* was assigned to the student numbers in all the EXCEL files. The data in the EXCEL files were then imported into SAS.

The subject code 381311 for the first semester and subject code 381321 for the second semester were used for the Mathematical Statistics course. The subject codes 172315 and 381315 for the first semester, and subject codes 172325 and 381325 for the second semester were used for the Applied Statistics course. The reason why Applied Statistics had two codes for each semester was that subject codes 172315 and 172325 were used from 1990 until 1996. After 1996, the subject codes were changed from 172315 and 172325 to 381315 and 381325 respectively. The following information was requested on students who did the above subjects: student number, student surname, student initials, third-year academic year, third-year Statistics exam mark, third-year Statistics supplementary exam mark, third-year Statistics exam comment and third-year Statistics supplementary exam comment.

The Grade 12 data for all the students who did at least one semester of third-year Statistics were requested as follows: year matriculated, Grade 12 exemption, Grade 12 aggregate (average), Grade 12 mathematics grades, Grade 12 mathematics symbols. The following academic year-level results were requested: final undergraduate academic year at UWC, degree code, degree name, academic year-level results. For these students, the following personal data were requested: sex of the student, race of the student, home language of the student and date of birth in yyyymmdd format. The year of first enrolment (variable name *begyear*) was extracted from the student number.

3.9 Measurements

In the study, the outcome of interest was successful throughput. Successful throughput meant that the student should have completed his/her undergraduate studies in three consecutive years from the year of first-time enrolment. Students who took more than three years to complete their studies or dropped out were considered as unsuccessful throughput students.

The throughput response indicator variable was called *through*. The categorical random variable *through* is a nominal scale measurement with discrete data. A '**1**' indicates that a student successfully completed his/her studies in three consecutive years, and a '**0**' indicates that a student did not complete his/her studies in three consecutive years. The category labels for the variable *through* were defined as 1="**THROUGHPUT**" and 0="**NON-THROUGHPUT**".

The following variables were needed to determine the throughput response variable: the variable *endyear* indicated one of the following events - the final year the student completed his/her undergraduate study at UWC or the year the student dropped out at UWC or the student is still currently in the system at UWC in 2004. All years were recorded as four digits, for example, 1997, in the study. The variable *begyear* indicated the year the student first enrolled at UWC. A new variable, *compl,* was computed by subtracting *begyear* from *endyear*. This new variable, *compl,* gives the number of years a student studied at UWC. The values for *compl* are discrete. If the value in *compl* was equal to '**3**', then the student finished his/her studies in the prescribed time of three years.

The demographic variables which describe the students in the population were the variable *gender*, indicating the sex of the students, and the variable *race*, indicating which race group a student belonged to. The categories were: '**COLOURED**', '**AFRICAN**', '**INDIAN**' and '**WHITE**'. The variable *homelang* indicated the language the student spoke at home. The categories were '**AFRIKAANS**', '**ENG & AFR**', '**ENGLISH**', '**NORTH SOTHO**', '**SOUTH SOTHO**', '**SWATI**', '**TSONGA**', '**TSWANA**', '**VENDA**', '**XHOSA**', '**ZULU**' and '**OTHER**'. A category was created for people who spoke both English and Afrikaans at home since Afrikaans was one of the two official languages during apartheid and both languages were spoken in many homes. It was assumed that the Africans only spoke their African languages at home and not a mixture of, for example, Xhosa and English.

School background information, such as Grade 12 aggregate and Grade 12 mathematics, was used in the study. The variable *agg_sym* indicated the Grade 12 aggregate (average) symbol. The categories were: '**A**', '**B**', '**C**', '**D**', '**E**' and '**F**'. The variable *math_grd* indicated the Grade 12 mathematics grade category of higher grade or standard grade. The categories were: '**H**' for higher grade and '**S**' for standard grade. The variable *math_sym* indicated the Grade 12 mathematics symbol. The categories were: '**A**', '**B**', '**C**', '**D**', '**E**' and '**F**'. As the Grade 12 mathematics symbols are related to higher grade and standard grade, a common scale was needed for comparison purposes. The variable *common1* was used to transform the Grade 12 mathematics grades and Grade 12 mathematics symbols to a common scale, namely, that an '**A**' on standard grade is equivalent to a '**B**' on higher grade; a '**B**' on standard grade is equivalent to a '**C**' on higher grade, and so forth. The categories were: '**A**', '**B**', '**C**', '**D**', '**E**', '**F**' and '**G**'.

A logistic regression model was built using the following predictor variables. The variable *gender* was included. The variable *race* was categorized as follows: all the African students were grouped into a category "AFRICAN" and the Coloured, White and Indian students were categorized as "NON-AFRICAN", which formed the new predictor variable *african*. The category labels were: 1="**AFRICAN**" and 0="**NON-AFRICAN**". This categorical random variable *african* is a nominal scaled measurement which was included in the modelling procedure.

The predictor variable *english* was created with all the English home language speaking students and the English and Afrikaans (speaking both languages) home language speaking students in one group versus all the other home language speaking students into the alternative group. The category labels were: 1="**ENGLISH**" and 0="**NON-ENGLISH**". The categorical random variable *english* is a nominal scaled measurement.

The predictor variable *agg_grp* was created using the academic background of a student entering UWC. The student either had a Grade 12 aggregate symbol of 60% and above (that is C and above) or below 60% (D and below). The category labels were: 1="**60% AND ABOVE**" (A, B and C) and 0="**BELOW 60%**" (D, E and F). The categorical random variable *agg_grp* is an ordinal scaled measurement.

The predictor variable *math_grp* was created using the *common1* variable, which was divided into two groups. The student either had a Grade 12 mathematics symbol of 60% and above (that is C and above) or below 60% (D and below). The category labels were: 1="**60% AND ABOVE**" (A, B and C) and 0="**BELOW 60%**" (D, E, F and G). The categorical random variable *math_grp* is an ordinal scaled measurement.

The predictor variable *immediate* indicated that the student had either enrolled at UWC immediately after leaving school (if the variable *imed_yrs* is equal to one or zero) or after some years (if the variable *imed_yrs* is more than one). The variable *imed_yrs* was the number of years between school and entrance into university. If *imed_yrs* was equal to zero, it meant that the student had matriculated in the same year he/she enrolled at UWC. For example, the student had failed a subject in Grade 12, written a supplementary exam the following year, and then matriculated while enrolled at UWC in that same year. The variable *imed_yrs* was calculated by subtracting the year the student matriculated (variable *matyear*) from the year the student enrolled for the first time (variable *begyear*). The values of variable *imed_yrs* are discrete. In the variable *imed_yrs*, '**1**' meant a student had entered UWC immediately after school; '**2**' meant a student had entered university after one year, and so on. The categories of variable *immediate* were: 1= "**DIRECTLY AFTER SCHOOL**" and 0= "**NOT DIRECTLY AFTER SCHOOL**". The categorical random variable *immediate* is a nominal scaled measurement.

The years of first registration were grouped into two groups, namely: pre-democratic versus post-democratic election years. In this study, the pre-democratic election years were from 1975 to 1994, and the post-democratic election years were from 1995 to 2001.The predictor variable *year_cov* indicates pre-democratic election years and post-democratic election years. The categorical labels were: 1="**POST-ELECTION YEARS**" and 0="**PRE-ELECTION YEARS**". The categorical random variable *year_*cov is an ordinal scaled measurement which was included in the modelling procedure as a covariate. See Table 3.1 for an overview of the variables in the study.

3.10 Limitations of the study

All academic years follow a calendar year. A student who finished in three and half years was recorded as finishing in four years. If a student repeated a subject, the highest mark obtained over all the years the student repeated the subject was recorded. Students who registered for both semesters but did not attend class, did not write examinations, and had no course mark for either semester were excluded from the study. Verification of the data was not required as it was requested from the UWC student database, which is assumed to be correct. There are cases where the information concerning Grade 12 results are missing, for example the Grade 12 aggregate. The study does not investigate the throughput of students who major in Statistics because the sample would then become too small for modelling purposes.

Table 3.1 **Table of variable names**

Variable name	Variable description	Variable created from
studnum	Student number	Original
Surname	Student surname	Original
init	Student initials	Original
year	Third-year academic year	Original
exam	Third-year Statistics exam mark	Original
sup	Third-year Statistics supplementary exam mark	Original
exam_cmt	Third-year Statistics exam comment	Original
*sup_*cmt	Third-year Statistics supplementary exam comment	Original
matyear	Year matriculated	Original
exemp	Grade 12 exemption	Original
agg_sym	Grade 12 aggregate (average)	Original
math_grd	Grade 12 mathematics grade	Original
math_sym	Grade 12 mathematics symbol	Original
endyear	Final undergraduate academic year at UWC	Original
degcode	Degree code	Original
degname	Degree name	Original
result	Academic year level results	Original
gender	Sex of the student	Original
race	Race of the student	Original
homelang	Home language of the student	Original
dobirth	Date of birth in yyyymmdd format	Original
begyear	Year of first enrolment	Derived from *studnum*
compl	Number of years at UWC	Derived from *endyear* minus *begyear*
through	Successful throughput	Derived from *compl*
common1	Common scale mathematics	Derived from *math_grd* and *math_sym*
african	African race category	Derived from *race*
english	Home language category	Derived from *homelang*
agg_grp	Grade 12 aggregate grouping	Derived from *agg_sym*
math_grp	Grade 12 mathematics based on common scale grouping	Derived from *common1*
immediate	Directly enters UWC after school grouping	Derived from *imed_yrs*
imed_yrs	Number of years after school before entering UWC	Derived from *begyear* minus *matyear*
year_cov	Covariate year influence	Derived from *begyear*

University *of the* Western Cape Library

UWC LIBRARY
empowering our clients

Website: http://www.uwc.ac.za/library

Library Hours:

Monday, Tuesday and Thursday:
08h20-22h00

Wednesday:
09h20-22h00

Friday:
08h20-16h30

Saturday:
09h00-17h00

Extended hours during Exam times and vacation periods

Private Bag X17,
Bellville, 7535
Cape Town, SA
Tel: 021 959 2209
Fax: 021 959 2659

UNIVERSITY *of the* WESTERN CAPE

Services and Collections:

LEVEL 1
Auditorium
LEVEL 2
Study Hall
LEVEL 4
Interlibrary Loan
LEVEL 5
Circulation desk
Reserve Collection
Photocopying
LEVEL 6
Information Services
(Faculty Librarians)
Theses, Indexes and
Abstracts
LEVEL 7
Multimedia Centre
Open Shelves
LEVEL 8
Open Shelves
LEVEL 9
Law Collection
LEVEL 10
Periodicals
LEVEL 11
Open Shelves
LEVEL 12, 13
Special Collections
Government publications
UWC publications
LEVEL 13
Postgraduate Resource Centre
LEVEL 14
Open Shelves

Electronic Resources
The Library provides access, via its website, to online databases, ejournals and other web resources
http://www.uwc.ac.za/library

Library Information Literacy Programme
The Library offers the following user education for students and staff:

- Databases Training
- Basic Search Skills Training
- Workshops for Postgraduates
- Workshops for Academic Staff
- Introduction to eResources.

and lots more!

You may contact a Faculty Librarian or the eResources and Training Librarian to book a place in one of our sessions. Watch the library's website and notice boards for more information.

3.11 Data analysis

The data requested were imported from a text file into Microsoft EXCEL. The SAS software was used to transform data into a format ready for analysis. The data were analyzed using descriptive statistics, frequencies and cross tabulations. Associations between nominal scaled variables were tested using Chi-Square or Fisher's Exact Tests. Models were built using logistic regression and decision trees.

In the next chapter, the analysis of the results will be reported.

Chapter 4

Analyses and results

4.1 Demographic background of students

The study consisted of 409 students who enrolled from 1975 to 2001 and who completed at least one semester of Statistics at third-year level (see Table A7, Appendix A). The study was comprised of 117 males (43.28%) and 232 females (56.72%) (see Table A1 in Appendix A). There were 230 African students (56.23%), 156 Coloured students (38.14%), 22 Indian students (5.38%) and one White student (0.24%) (see Table A2 in Appendix A). The most common home language spoken by students was Xhosa (32.52%), followed by English and Afrikaans (22.49 + 8.56= 31.05%). The following languages were spoken the least, in decreasing order – Venda (1.96%), Tonga (1.47%) and Swati (1.22%) (see Table 4.1 and Table A3 in Appendix A).

Table 4.1 **List of home languages**

Home language	Frequency	Percentage
Xhosa	133	32.52
English	92	22.49
Afrikaans	51	12.47
English and Afrikaans (both)	35	8.56
Tswana	30	7.33
South Sotho	17	4.16
Zulu	12	2.93
North Sotho	12	2.93
Other	8	1.96
Venda	8	1.96
Tsonga	6	1.47
Swati	5	1.22

4.2 Third-year Statistics course

In the study, 205 (50.12%) students registered for the Mathematical Statistics course and 204 (49.88%) for the Applied Statistics course (see Table A21 in Appendix A). Of the 409 students, 361 (88.26%) passed both semesters and majored in Statistics (see Table A23 in Appendix A). The students who did not major in Statistics (11.74%) either failed both semesters (3.18%) or passed only one semester of third-year Statistics (8.56%) (see Table A22 in Appendix A).

4.3 Number of years to complete studies

More than 50% of the students (29.83 + 24.45 = 54.28%) took between four and five years to complete their studies (see Table A9 in Appendix A). The average number of years they took to complete their studies was five years; the median was four years (see Table A10 in Appendix A).

4.4 Grade 12 results

4.4.1 Aggregate

Most students entered UWC with a 'D' aggregate (46.53%) (see Table 4.2 and Table A12 in Appendix A). Of 404 students, 134 (33.18%) students achieved an aggregate of 60% and above. From Table 4.2, it can be seen that 82 (20.3%) students entered UWC with an aggregate below a D (less than 50%).

Table 4.2 **Grade 12 aggregates**

Symbol	Frequency	Percentage
A	2	0.50
B	26	6.44
C	106	26.24
D	188	46.53
E	78	19.31
F	4	0.99

Note: Five missing values.

4.4.2 Mathematics

In the study, 198 students (48.89%) had taken mathematics on the higher grade, and 207 students (51.11%) had completed Grade 12 mathematics on the standard grade (note: four missing values) (see Table A14 in Appendix A). The majority of students entered UWC with an 'E' symbol in mathematics on the higher grade or a 'D' symbol on the standard grade (see Table 4.3). There were 32 students who entered UWC with symbols less than the requirement stipulated in the Science Faculty yearbook. Only 4 students had an 'A' symbol on the higher grade (see Table 4.3). The common scale was created for comparison purposes between the higher grade and the standard grade. An 'A' on the standard grade was set equivalent to a 'B' on the higher grade. There were 27 students who had an 'A' on the standard grade. These 27 students plus the 17 students with 'B' symbols on the higher grade add up to 44 students on the common scale. From Table 4.4, we can see that the majority of students (38.71%) had an 'E' symbol on the common scale.

Table 4.3 **Mathematics symbols**

Symbol	Higher grade count	Standard grade count
A	4	27
B	17	38
C	26	45
D	58	64
E	92	24
F	0	8

Note: Six missing values.

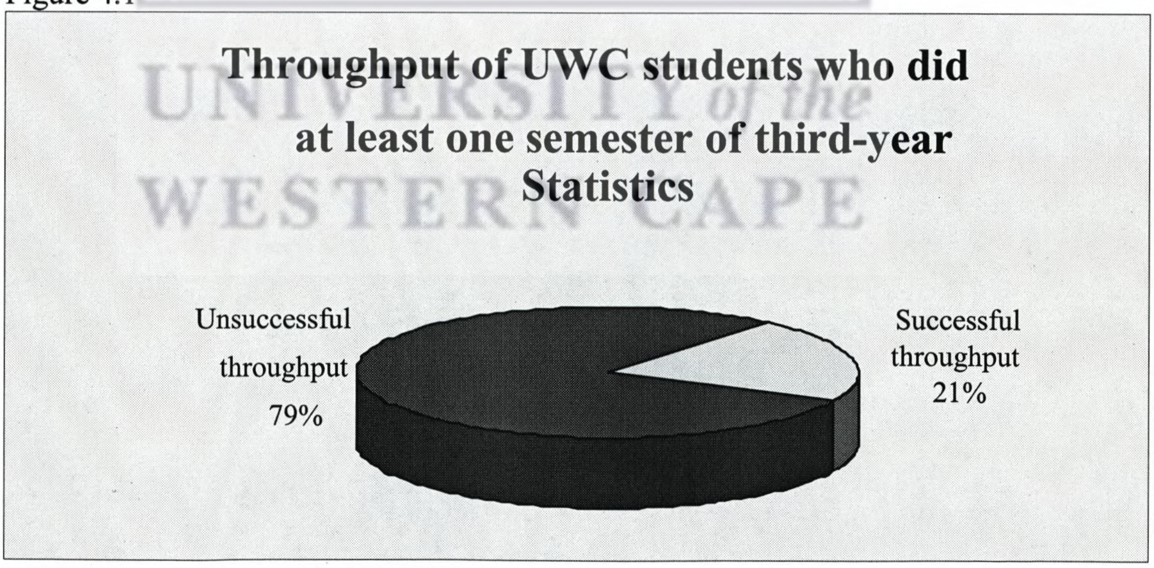
Table 4.4 **Common scale symbols**

Common scale	Frequency	Percentage
A	4	0.99
B	44	10.92
C	64	15.88
D	103	25.56
E	156	38.71
F	24	5.96
G	8	1.99

4.5 Response variable

In the study of 409 students, 86 students (21.03%) finished their studies successfully in the prescribed time of three years (see Table A11 in Appendix A). The other 323 students (78.97%) either took more than three years to finish their studies or dropped out or are still currently registered (see Figure 4.1).

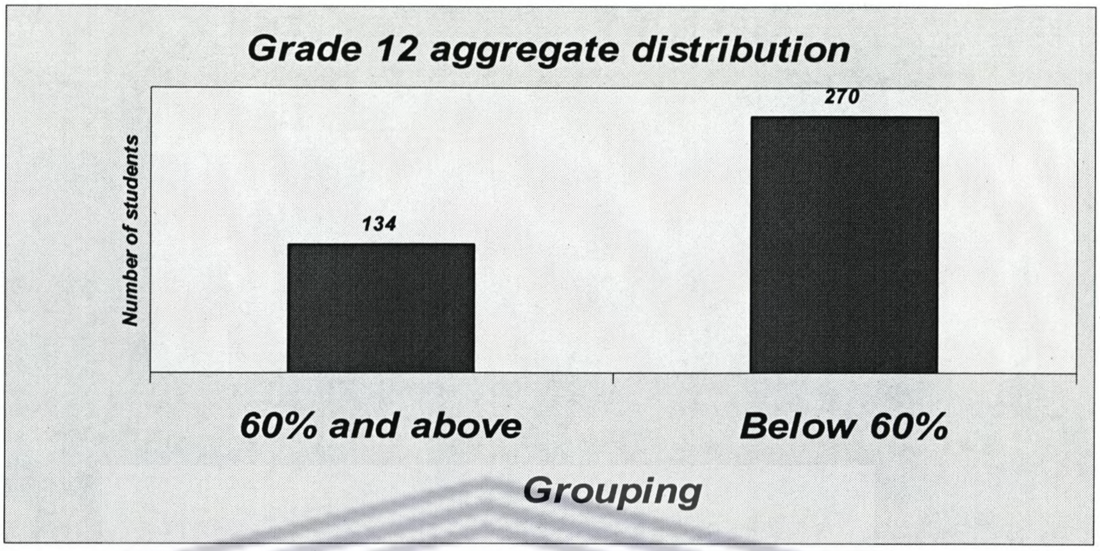
Figure 4.1



4.6 Predictor variables

In this section, the distribution of the predictors will be described (see Table 4.5). The predictor variable *african* had 230 African students (56.23%) and 179 non-African students (43.77%) (see Table A4 in Appendix A). There were 127 English home language speaking students (31.05%) and 282 non-English home language speaking students (68.95%) (see Table A5 in Appendix A). From Figure 4.2, it can be seen there were 134 students (33.17%) who had a Grade 12 aggregate of 60% and above, and 270 students (66.83%) who had a Grade 12 aggregate below 60% (note: 5 missing values) (see Table A13 in Appendix A). There were 112 students (27.79%) who had a common-scale Grade 12 mathematics symbol of 60% and above, and 291 students (72.21%) who had a Grade 12 mathematics symbol below 60% (note: 6 missing values) (see Table A18 in Appendix A). There were 238 students (58.19%) who entered UWC immediately after school, and 171 students (41.81%) who had a break of some years before they enrolled at UWC (see Table A20 in Appendix A). There were 213 students (52.08%) who enrolled at UWC for the first time after the 1994 democratic election, and 196 students (47.92%) who were enrolled for the first time before the 1994 democratic election (see Table A8 in Appendix A).

Figure 4.2



4.7 Throughput associations (refer to Table 4.6)

4.7.1 Gender

The rate of successful throughput, given it was a female, was 30/177 (16.95%) compared to the rate of successful throughput, given that it was a male, was 56/232 (24.14%). The throughput among gender did not differ at a 5% level of significance (Chi-square test, $\chi^2 = 3.1246$; p-value=0.0771) (see Table A24 in Appendix A).

4.7.2 African

The probability of successful throughput, given an African, was 40/230 (17.39%), and the probability of successful throughput, given a non-African, was 46/179 (25.70%). Non-African students had a significantly higher throughput rate compared to African students (Chi-square test, $\chi^2 = 4.1831$; p-value= 0.0408) (see Table A25 in Appendix A).

4.7.3 English

Only 36 of 127 English-speaking students (28.35%) were successful in completing their studies in the prescribed time of three years, compared to 50 of 282 non-English-speaking students (17.73%) who had completed their studies in the prescribed time of three years. The English-speaking students had a significantly higher throughput rate compared to the non-English students (Chi-square test, $\chi^2 = 5.9428$; p-value = 0.0148) (see Table A26 in Appendix A).

4.7.4 Aggregate

The probability of successful throughput, given the students' aggregate symbol was 60% and above, was 43/134 (32.09%) versus the probability of successful throughput, given the students' symbol was below 60%, was 43/270 (15.93%). The students with an aggregate symbol of 60% and above had a significantly higher throughput rate than those students who had an aggregate symbol below 60% (Chi-square test, $\chi^2 = 13.9637$; p-value= 0.0002) (see Table A27 in Appendix A).

4.7.5 Mathematics

In the 60%-and-above group for mathematics, the rate of successful throughput was 32/112 (28.57%) and the rate for the below-60% group was 53/291 (18.21%). The throughput rate of students whose symbols were 60% and above for mathematics was significantly higher than those who had below 60% for mathematics on a common scale (Chi-square test, $\chi^2 = 5.2138$; p-value= 0.0224) (see Table A28 in Appendix A).

4.7.6 Immediately

The rate of successful throughput of the student who entered UWC immediately after school was 44 out of 238 (18.49%) compared to the rate of successful throughput of those who did not enter UWC immediately after school, which was 42 out of 171 (24.56%). The break between school and university did not significantly influence the throughput rate (Chi-square test, χ^2 = 2.2108; p-value= 0.1370) (see Table A29 in Appendix A).

4.7.7 Year covariate

The probability of successful throughput of those who registered after the 1994 election was 61/213 (28.64%) compared to the probability before the 1994 election, which was 25/196 (12.76%). The throughput rate increased significantly after the 1994 elections (Chi-square test, χ^2 = 15.5076; p-value = < 0.0001) (see Table A30 in Appendix A).

Table 4.5 **Throughput versus predictor associations**

Predictor	Chi-square p-value	Conclusion
Year covariate	< 0.0001	Significant
Aggregate	0.0002	Significant
English	0.0148	Significant
Mathematics	0.0224	Significant
African	0.0408	Significant
Gender	0.0771	Non-significant
Immediately	0.1370	Non-significant

Note: The conclusion column is based upon a significance level of 5%.

Table 4.5 is a summary of the Chi-square p-values of all the predictors. As can be seen, the highest significant predictor with successful throughput was the Year covariate, followed by the Aggregate predictor. The predictors Gender and Immediately were not significantly related to successful throughput. The table below gives a global view of throughput cross tabulated with all the predictors.

Table 4.6 **Probability of throughput given predictor**

(refer to Tables A24 to Table A30 in Appendix A)

Predictor Total frequency (Percentage)	Throughput 86 (21.03%)	Non-throughput 323 (78.97%)	Chi- square p-value
Female = 177 (43.28%)	30 (16.95%)	147 (83.05%)	0.0771
Male = 232 (56.72%)	56 (24.14%)	176 (75.86%)	
African = 230 (56.23%)	40 (17.39%)	190 (82.61%)	0.0408 *
Non-African = 179 (43.77%)	46 (25.70%)	133 (74.30%)	
Cramer's V = - 0.1011			
English = 127 (31.05%)	36 (28.35%)	91 (71.65%)	0.0148 *
Non-English = 282 (68.95%)	50 (17.73%)	232 (82.27%)	
Cramer's V = 0.1205			
60% and above aggregate = 134 (33.17%)	43 (32.09%)	91 (67.91%)	0.0002 * *
Below 60% aggregate = 270 (66.83%)	43 (15.93%)	227 (84.07%)	
Cramer's V = 0.1859			
60% and above mathematics = 112 (27.79%)	32 (28.57%)	80 (71.43%)	0.0224 *
Below 60% mathematics = 291 (72.21%)	53 (18.21%)	238 (81.79%)	
Cramer's V = 0.1137			
Directly after school = 238 (58.19%)	44 (18.49%)	194 (81.51%)	0.1370
Not Directly after school = 171 (41.81%)	42 (24.56%)	129 (75.44%)	
Post-election = 213 (52.08%)	61 (28.64%)	152 (71.36%)	<0.0001 * *
Pre-election = 196 (47.92%)	25 (12.76%)	171 (87.24%)	
Cramer's V = 0.1947			

Note: 1. All the percentages in brackets are row percentages.
2. * Significant at a 5% level.
3. ** Significant at a 1% level.

4.8 Predictor associations

4.8.1 Gender

The Gender predictor variable was highly significant with the African predictor variable (Chi-square test, χ^2 = 10.9705; p-value=0.0009) (see Table A31 in Appendix A). From Table A31, it can be seen that there were more black female students.

A significant association exists between Gender and Aggregate (Chi-square test, χ^2 = 4.5884; p-value=0.0322) (see Table A33 in Appendix A). There were 143 males (62.45%) and 127 female (72.57%) who had an aggregate below 60%. There were 86 males (37.55%) and 48 females (27.43%) who had an aggregate of 60% and above.

Gender with the Mathematics predictor was highly significantly associated (Chi-square test, χ^2 = 11.1790; p-value=0.0008) (see Table A34 in Appendix A). In the category of 60% and above, there were more males (34.36%) than females (19.32%).

The gender difference between those who entered UWC immediately after school and those who did not was significant (Chi-square test, χ^2 = 9.2142; p-value=0.0024) (see Table A35 in Appendix A). More females (66.67%) than males (51.72%) enrolled at UWC directly after school.

4.8.2 African

A highly significant difference between the African and the non-African who speak the English language at home can be seen (Chi-square test, $\chi^2 = 223.605$; p-value= < 0.0001). (see Table A37 in Appendix A). There are only two non-Africans who speak English at home.

The Aggregate and African predictors are highly significantly associated with each other (Chi-square test, $\chi^2 = 94.2142$; p-value= < 0.0001) (see Table A38 in Appendix A). Most of the African students in the study attained an aggregate below 60% (196 out 404 students). There were 105 non-African students who had an aggregate of 60% and above compared to only 29 African students.

Mathematics was also significantly associated with the African predictor (Chi-square test, $\chi^2 = 45.5008$; p-value= < 0.0001) (see Table A39 in Appendix A). The majority of African students (194 out of 403 students) had a mathematics result on the common scale below 60%.

The difference between African and non-African students entering UWC immediately after school was highly significant (Chi-square test, $\chi^2 = 12.9939$; p-value= 0.0003) (see Table A40 in Appendix A). More African students (49.57%) than non-African students (31.84%) did not enter UWC directly after school. Sixty-eight percent of non-African and 50.43% African students entered UWC immediately after school.

4.8.3 English

The English predictor with Aggregate was highly significantly associated (Chi-square test, χ^2 = 40.2588; p-value= < 0.0001) (see Table A42 in Appendix A). More non-English home language speaking students (213 out of 404 students) entered UWC with an aggregate below 60%.

The difference between English and non-English home language speaking students, when comparing their mathematics results on a common scale, was significant (Chi-square test, χ^2 = 8.2615; p-value= 0.0040) (see Table A43 in Appendix A). More non-English home language speaking students (212 out of 403) had results below 60% for mathematics on the common scale.

4.8.4 Aggregate and mathematics

The Aggregate and Mathematics predictors were highly significantly related to each other (Chi-square test, χ^2 = 55.6667; p-value = < 0.0001) (see Table A46 in Appendix A). More than fifty percent (222 out of 398) of the students in the study had an aggregate below 60% and were in the below-60% group for mathematics. In the below-60% aggregate group, there were five times more students who had below 60% for mathematics compared to the students who had 60% and above on the common scale for mathematics. In the 60%-and-above aggregate group, there was not much difference between those who had a mathematics result below 60% compared to those who had 60% and above.

4.8.5 Immediately and Year covariate

The association between the predictor Immediate and the Year covariate was significant (Chi-square test, $\chi^2 = 4.0702$; p-value = 0.0436) (see Table A51 in Appendix A). In the era after the 1994 elections, 134 students enrolled directly after school at UWC, compared to the 104 students who entered in the era before the 1994 elections. The enrolment of students not entering UWC directly after school, dropped from 92 students (in the era before 1994) to 79 students (in the era after 1994). Table 4.7 is a summary of all the predictor associations.

Table 4.7 **Predictor associations**
(refer to Table A31 to Table A51 in Appendix A)

Chi-square p-value	Gender	African	English	Aggregate	Mathematics	Immediately	Year covariate
Gender		0.0009 **	0.1333	0.0322 *	0.0008 **	0.0024 **	0.0780
African			<0.0001 **	<0.0001 **	<0.0001 **	0.0003 **	0.1010
English				<0.0001 **	0.0040 **	0.0487	0.5019
Aggregate					<0.0001 **	0.6095	0.4781
Mathematics						0.1994	0.1089
Immediately							0.0436 *
Year covariate							

Note: 1. * Significant at a 5% level.
2. ** Significant at a 1% level.

Table 4.7 is a summary of Chi-square p-values of the predictor associations. The predictor Gender was significantly associated with the predictors African, Aggregate, Mathematics and Immediately. The predictor African was significantly associated with the predictors English, Aggregate, Mathematics and Immediately. The predictor English was significantly associated with the predictors Aggregate, Mathematics and Immediately. The predictor Aggregate was highly significantly associated with Mathematics. The predictor Immediately and the Year covariate were highly significantly associated with each other.

4.9 Logistic regressions of throughput - single predictors

In the next section, a logistic regression model for each predictor variable was built. Each model was evaluated by the percentage observations correctly predicted by the model. All models were evaluated at a probability threshold of 0.22 for comparison purposes.

4.9.1 Gender model (refer to Table 4.8 and Table B1 in Appendix B)

The logistic regression model for throughput using Gender as a predictor was:

$$\ln [\text{odds of throughput given Gender}] = \ln [\frac{p}{1-p}]$$

$$= - 1.3671 - 0.2220 * gender$$

Taking the exponential both sides in the above equation, we get the odds:

$$\frac{p}{1-p} = e^{(- 1.3671 - 0.2220 * gender)} = e^{(- 1.3671)} * e^{(- 0.2220*gender)}.$$

The odds of successful throughput, given it was a female ($gender = 1$), was

$e^{(- 1.3671)} * e^{(- 0.2220*1)} =0.2548 * 0.8009 = 0.204$ (i.e. 30/147 from Table 4.8),

and the odds of successful throughput, given that it was male ($gender = -1$), was

$e^{(- 1.3671)} * e^{(- 0.2220*-1)} =0.2548 * 1.2486 = 0.318$ (i.e. 56 / 176 from Table 4.8).

Comparing the above two odds, we see that the odds of successful throughput, given a male, was higher.

Making p the subject of the formula, we get the estimated probability $p = (1+ e^{-(- 1.3671 - 0.2220 * gender)})^{-1}$. Thus, the estimated probability of successful throughput, given it was a female ($gender = 1$), was

$(1 + e^{-(- 1.3671 - 0.2220 * 1)})^{-1} = (1 + e^{-(-1.5891)})^{-1} = (1 + 4.8993)^{-1} = (5.8993)^{-1} = 0.1695,$

and the estimated probability of successful throughput, given it was

a male (*gender* = -1), was

$(1 + e^{-(-1.3671 - 0.2220 * -1)})^{-1} = (1 + e^{-(-1.1451)})^{-1} = (1 + 3.1426)^{-1} = (4.1428)^{-1} = 0.2414$

The above two estimated probabilities correspond to the row percentages in Table 4.6. Comparing the above two estimated probabilities, we see the estimated probability of successful throughput for a male was higher.

The Gender model correctly predicted 49.6% of the observations at a probability level of 0.22. For a probability level of 0.16, the model correctly predicted only 21% of the observations.

4.9.2 African model (refer to Table 4.8 and Table B2 in Appendix B)

The logistic regression model for throughput using African as a predictor was:

ln [odds of throughput given African] = $\ln [\frac{p}{1-p}]$

$= -1.0616 - 0.4964 * african.$

Taking the exponential both sides in the above equation we get the odds:

$\frac{p}{1-p} = e^{(-1.0616 - 0.4964 * african)} = e^{(-1.0616)} * e^{(-0.4964 * african)}.$

The odds of successful throughput, given it was an African (*african* = 1), was $e^{(-1.0616)} * e^{(-0.4964*1)} = 0.3459 * 0.6087 = 0.211$ (i.e. 40/190 from Table 4.8) and the odds of successful throughput, given that it was non-African (*african* = 0), was $e^{(-1.0616)} * e^{(-0.4964*0)} = 0.3459$ (i.e. 46 / 133 from Table 4.8).

Comparing the above two odds, we see that the odds of successful throughput, given a non-African, was higher.

Making p the subject of the formula, we get the estimated probability

$p = (1 + e^{-(-1.0616 - 0.4964 * african)})^{-1}$.

Thus, the estimated probability of successful throughput, given it was an African (*african* = 1), was

$(1 + e^{-(-1.0616 - 0.4964 * 1)})^{-1} = (1 + e^{-(-1.5584)})^{-1} = (1 + 4.7512)^{-1} = (5.7512)^{-1}$

$= 0.1739$

and the estimated probability of successful throughput, given it was a non-African (*african* = 0), was

$(1 + e^{-(-1.0616 - 0.4964 * 0)})^{-1} = (1 + e^{-(-1.0617)})^{-1} = (1 + 2.8913)^{-1} = (3.8913)^{-1}$

$= 0.2570$.

The above two estimated probabilities correspond to the row percentages in Table 4.6. Comparing the above two estimated probabilities, we see the estimated probability of successful throughput for a non-African student was higher.

The African model correctly predicted 57.7% of the observations at a probability level of 0.22. For a probability level of 0.16, the model correctly predicted only 21% of the observations.

4.9.3 English model (refer to Table 4.8 and Table B3 in Appendix B)

The logistic regression model for throughput using English as a predictor was:

ln [odds of throughput given English] = ln $[\frac{p}{1-p}]$

$= - 1.5347 + 0.6074 * english.$

Taking the exponential both sides in the above equation, we get the odds:

$$\frac{p}{1-p} = e^{(-1.5347 + 0.6074 * english)} = e^{(-1.5347)} * e^{(0.6074*english)}.$$

The odds of successful throughput, given it was an English (*english*=1), was

$e^{(-1.5347)} * e^{(0.6074*1)}$ =0.2155 * 1.8357 = 0.396 (i.e. 36/91 from Table 4.8),

and the odds of successful throughput, given that it was a non-English (*english*=0), was $e^{(-1.5347)} * e^{(0.6074*1)}$ =0.2155 (i.e. 50/232 from Table 4.8). Comparing the above two odds, we see that the odds of successful throughput for an English speaking student was higher.

Making p the subject of the formula, we get the estimated probability

$$p = (1+ e^{-(-1.5347 + 0.6074 * english)})^{-1}.$$

Thus, the estimated probability of successful throughput, given it was an English (*english*=1), was

$$(1 + e^{-(-1.5347 + 0.6074 * 1)})^{-1} = (1 + e^{-(-0.9273)})^{-1} = (1 + 2.5277)^{-1} = (3.5277)^{-1}$$

= 0.2835,

and the estimated probability of successful throughput, given that it was a non-English (*english*=0), was

$$(1 + e^{-(-1.5347 + 0.6074 * 0)})^{-1} = (1 + e^{-(-1.5347)})^{-1} = (1 + 4.64)^{-1} = (5.64)^{-1} = 0.1773.$$

The above two estimated probabilities correspond to the row percentages in Table 4.6. Comparing the above two estimated probabilities, we see the estimated probability of successful throughput for an English speaking student was higher.

The English model correctly predicted 65.5% of the observations at a probability level of 0.22. For a probability level of 0.16, the model correctly predicted only 21% of the observations.

4.9.4 Aggregate model (refer to Table 4.8 and Table B4 in Appendix B)

The logistic regression model for throughput using Aggregate as a predictor was:

ln [odds of throughput given Aggregate] = ln [$\frac{p}{1-p}$]

$$= - 1.6637 + 0.9141 * agg\_grp$$

Taking the exponential both sides in the above equation we get the odds:

$$\frac{p}{1-p} = e^{(- 1.6637 + 0.9141 * agg\_grp)} = e^{(- 1.6637)} * e^{(0.9141 * agg\_grp)}$$

The odds of successful throughput, given an aggregate of 60% and above (*agg_grp* = 1), was $e^{(- 1.6637)} * e^{(0.9141 * 1)}$ =0.1894 * 2.4945 = 0.473 (i.e. 43/91 from Table 4.8), and the odds of successful throughput, given an aggregate below 60% (*agg_grp* = 0), was $e^{(- 1.6637)} * e^{(0.9141 * 0)}$ =0.1894 (i.e. 43 / 227 from Table 4.8).

Comparing the above two odds, we see that the odds of successful throughput, given that the Aggregate was 60% and above, was higher.

Making p the subject of the formula, we get the estimated probability

$p = (1 + e^{-(-1.6637 + 0.9141 * agg\_grp)})^{-1}$.

Thus, the estimated probability of successful throughput, given an aggregate of 60% and above (*agg_grp* = 1), was

$(1 + e^{-(-1.6637 + 0.9141 *1)})^{-1} = (1 + e^{-(-0.7497)})^{-1} = (1 + 2.1164)^{-1} = (3.1164)^{-1} = 0.3209$,

and the estimated probability of successful throughput, given an aggregate below 60% (*agg_grp* = 0), was

$(1 + e^{-(-1.6637 + 0.9141 *0)})^{-1} = (1 + e^{-(-1.6637)})^{-1} = (1 + 5.2788)^{-1} = (6.2788)^{-1}$

$= 0.1593$.

The above two estimated probabilities correspond to the row percentages in Table 4.6. Comparing the above two estimated probabilities, we observed that the estimated probability of successful throughput for an aggregate of 60% and above was higher.

The Aggregate model correctly predicted 66.8% of the observations at a probability level of 0.22. For a probability level of 0.14, the model correctly predicted only 21.3% of the observations.

4.9.5 Mathematics model (refer to Table 4.8 and Table B5 in Appendix B)

The logistic regression model for throughput using Mathematics as a predictor was:

$$\ln[\text{odds of throughput given Mathematics}] = \ln\left[\frac{p}{1-p}\right]$$

$$= -1.5020 + 0.5857 * math\_grp.$$

Taking the exponential both sides in the above equation, we get the odds:

$$\frac{p}{1-p} = e^{(-1.5020 + 0.5857 *math\_grp)} = e^{(-1.5020)} * e^{(0.5857 *math\_grp)}.$$

The odds of successful throughput, given mathematics was 60% and above, (*math_grp* = 1) was $e^{(-1.5020)} * e^{(0.5857 * 1)} = 0.2227 * 1.7962 = 0.4$ (i.e. 32/80 from Table 4.8), and the odds of successful throughput, given mathematics was below 60% (*math_grp* = 0), was $e^{(-1.5020)} * e^{(0.5857 * 0)} = 0.2227$ (i.e. 53 / 238 from Table 4.8). Comparing the above two odds, we see that the odds of successful throughput, given mathematics was 60% and above, was higher.

Making p the subject of the formula, we get the estimated probability

$p = (1 + e^{-(-1.5020 + 0.5857 * math\_grp)})^{-1}$.

Thus, the estimated probability of successful throughput, given mathematics was 60% and above (*math_grp* = 1), was

$(1 + e^{-(-1.5020 + 0.5857 * 1)})^{-1}$

$= (1 + e^{-(-0.9163)})^{-1} = (1 + 2.5)^{-1} = (3.5)^{-1} = 0.2857$,

and the estimated probability of successful throughput, given mathematics was below 60% (*math_grp* = 0), was

$(1 + e^{-(-1.5020 + 0.5857 * 0)})^{-1}$

$= (1 + e^{-(-1.5020)})^{-1} = (1 + 4.4907)^{-1} = (5.4907)^{-1} = 0.1821$

The above two estimated probabilities correspond to the row percentages in Table 4.6. Comparing the above two estimated probabilities, we observed that the estimated probability of successful throughput for mathematics at 60% and above was higher.

The Mathematics model correctly predicted 67% of the observations at a probability level of 0.22. For a probability level of 0.16, the model correctly predicted only 21.1% of the observations.

4.9.6 Immediately model (refer to Table 4.8 and Table B6 in Appendix B)

The logistic regression model for throughput using Immediately as a predictor was:

$$\ln[\text{odds of throughput given Immediately}] = \ln\left[\frac{p}{1-p}\right]$$

$$= -1.1221 - 0.3615 * immediate.$$

Taking the exponential both sides in the above equation, we get the odds:

$$\frac{p}{1-p} = e^{(-1.1221 - 0.3615 * immediate)} = e^{(-1.1221)} * e^{(-0.3615 * immediate)}.$$

The odds of successful throughput given entering UWC directly after school (*immediate* = 1) was $e^{(-1.1221)} * e^{(-0.3615 * 1)} = 0.3256 * 0.6966 = 0.227$ (i.e. 44/194 from Table 4.8), and the odds of successful throughput, given not entering UWC directly after school (*immediate* = 0), was $e^{(-1.1221)} * e^{(-0.3615 * 0)} = 0.3256$ (i.e. 42 / 129 from Table 4.8). Comparing the above two odds, we see that the odds of successful throughput, given not entering UWC directly after school, was higher.

Making p the subject of the formula, we get the estimated probability

$$p = (1 + e^{-(-1.1221 - 0.3615 * immediate)})^{-1}$$

Thus, the estimated probability of successful throughput,

given entering UWC directly after school (*immediate* = 1), was

$(1 + e^{-(-1.1221 - 0.3615 * 1)})^{-1} = (1 + e^{-(-1.4837)})^{-1} = (1 + 4.4092)^{-1} = (5.4092)^{-1}$

$= 0.1849$,

and the estimated probability of successful throughput, given not entering UWC directly after school (*immediate* = 0), was

$(1 + e^{-(-1.1221 - 0.3615 * 0)})^{-1} = (1 + e^{-(-1.1221)})^{-1} = (1 + 3.0713)^{-1} = (4.0713)^{-1}$

$= 0.2456$

The above two estimated probabilities correspond to the row percentages in Table 4.6. Comparing the above two estimated probabilities, we see the estimated probability of successful throughput for not entering UWC directly after school was higher.

The Immediately model correctly predicted 57.7% of the observations at a probability level of 0.22. For a probability level of 0.18, the model correctly predicted only 21% of the observations.

4.9.7 Year covariate model (refer to Table 4.8 and Table B7 in Appendix B)

The logistic regression model for throughput using the Year covariate as a predictor was: ln [odds of throughput given Year covariate] = $\ln [\frac{p}{1-p}]$

$$= -1.9228 + 1.0098 * year\_cov$$

Taking the exponential both sides in the above equation, we get the odds:

$$\frac{p}{1-p} = e^{(-1.9228 + 1.0098 * year\_cov)} = e^{(-1.9228)} * e^{(1.0098 * year\_cov)}.$$

The odds of successful throughput for enrolments after the 1994 election (*year_cov* =1) was

$e^{(-1.9228)} * e^{(1.0098 * 1)} = 0.1462 * 2.7451 = 0.401$ (i.e. 61/152 from Table 4.8), and the odds of successful throughput for enrolments before 1994 election (*year_cov* =0) was

$e^{(-1.9228)} * e^{(1.0098 * 0)} = 0.1462$ (i.e. 25 / 171 from Table 4.8). Comparing the above two odds, we see that the odds of successful throughput, given post-election, was higher. Making p the subject of the formula, we get the estimated probability $p = (1 + e^{-(-1.9228 + 1.0098 * year\_cov)})^{-1}$

Thus, the estimated probability of successful throughput for enrolments after 1994 election (*year_cov* =1) was $(1 + e^{-(-1.9228 + 1.0098 * 1)})^{-1} = (1 + e^{-(-0.9130)})^{-1}$

$= (1 + 2.4918)^{-1} = (3.4918)^{-1} = 0.2864$,

and the estimated probability of successful throughput for enrolments before the 1994 election (*year_cov* =0) was

$(1 + e^{-(-1.9228 + 1.0098 * 0)})^{-1} = (1 + e^{-(-1.9228)})^{-1} = (1 + 6.8401)^{-1} = (7.8401)^{-1} = 0.1275$

The above two estimated probabilities correspond to the row percentages in Table 4.6. Comparing the above two estimated probabilities, we observed that the estimated probability of successful throughput for enrolments after the 1994 election was higher.

The Year covariate model correctly predicted 56.7% of the observations at a probability level of 0.22. For a probability level of 0.12, the model correctly predicted only 21% of the observations. Table 4.8 for gives a summary of all the single predictor variables.

Table 4.8 **Throughput logistic regression models for the seven predictors**

(refer to Table B1 to Table B7 in Appendix B)

Predictor (value used to calculate ln odds from model)	Throughput (Frequencies)	Non-throughput (Frequencies)	Odds ratio (OR)	Logistic regression model
				Model = - 1.3671 – 0.2220 * *gender*
Female (= 1)	30	147	0.641	Model(*gender*=1) = - 1.5892= ln (30/147)
Male (= - 1)	56	176		Model(*gender*= -1) = - 1.1451 = ln 56/176)
	Odds=30/56	Odds=147/176		
				Model= - 1.0616 – 0.4964 * *african*
African (= 1)	40	190	0.609	Model(*african*=1) =- 1.5584 = ln (40/190)
Non-African (= 0)	46	133		Model(*african*=0) = - 1.0617 = ln (46/133)
	Odds=40/46	Odds=190/133		
				Model= - 1.5347 + 0.6074 * *english*
English (= 1)	36	91	1.836	Model(*english*=1) = - 0.9273 = ln (36/91)
Non-English (= 0)	50	232		Model(*english*=0) = - 1.5347 = ln (50/232)
	Odds=36/50	Odds=91/232		
				Model= - 1.6637 + 0.9141 * *agg_grp*
60% and above aggregate (= 1)	43	91	2.495	Model(*agg_grp* =1) = - 0.7497 = ln (43/91)
Below 60% aggregate (= 0)	43	227		Model(*agg_grp* =0) = - 1.6637 = ln (43/227)
	Odds=43/43	Odds=91/227		
				Model= - 1.5020 + 0.5857 *math_grp
60% and above mathematics (= 1)	32	80	1.796	Model(math_grp=1) = - 0.9163 = ln (32/80)
Below 60% mathematics (= 0)	53	238		Model(math_grp=0) = - 1.5020 = ln (53/238)
	Odds=32/53	Odds=80/238		
				Model= - 1.1221 – 0.3615 * immediate
Directly after school (= 1)	44	194	0.697	Model(immediate=1) = - 1.4837 = ln (44/194)
Not Directly after school (= 0)	42	129		Model(immediate=0) = - 1.1221 = ln (42/129)
	Odds=44/42	Odds=194/129		
				Model= - 1.9228 + 1.0098 * year_cov
Post-election (= 1)	61	152	2.745	Model(year_cov=1) = - 0.9130 = ln (61/152)
Pre-election (= 0)	25	171		Model(year_cov=0) = - 1.9228 = ln (25/171)
	Odds=61/25	Odds=152/171		

Note: 1. Males are equal to -1 because the variable *gender* is a string variable.
2. Odds Ratio (OR) is Odds of throughput divided by Odds of non-throughput.

After having evaluated the individual predictor models, the next step was to use either all or some of the predictors in one model. Three models were evaluated: the full model with all predictor variables, the full model without the Year covariate and the stepwise selection model.

4.10 Logistic regression of throughput - many predictors

4.10.1 Full logistic regression model

ln [odds of throughput given all the predictors] = $\ln \left[\frac{p}{1-p} \right]$

$= -2.5021 - 0.1942 * gender + 0.2054 * african + 0.5922 * english + 0.68869 * agg\_grp + 0.4032 * math\_grp - 0.4343 * immediate + 1.1449 * year\_cov$

From the p-values of the parameters, it was seen that only the intercept (< 0.0001), the aggregate (0.0256) and the year_cov (< 0.0001) were significant in the full model (see Table C1 Appendix C).

The full model correctly predicted 68.3% of the observations at a probability level of 0.22. For a probability level of 0.04, the model correctly predicted only 21.4% of the observations (see Table C1 in Appendix C). In the next model, the Year covariate was removed to evaluate the effect it had on the full model.

4.10.2 Full logistic regression model without the year covariate

ln [odds of throughput given all predictors without the Year covariate] = ln [$\frac{p}{1-p}$]

= -1.9869 – 0.186 * *gender* + 0.3801 * *african* + 0.6022 * *english* +

0.8117 * *agg_grp* + 0.2545 **math_grp* – 0.2899 **immediate*.

From the p-values of the parameters, it was seen that only the intercept (< 0.0001) and Aggregate (0.0071) were significant in the model (see Table D1 Appendix D).

The model correctly predicted 64.1% of the observations at a probability level of 0.22. For a probability level of 0.06, the model correctly predicted only 21.4% of the observations (see Table D1 in Appendix D). In the next model, the stepwise selection method was applied to select the best variables for the model.

4.10.3 Stepwise logistic regression model

ln [odds of throughput given aggregate and year_cov predictors] = ln [$\frac{p}{1-p}$]

= -2.2755 + 0.9374 * *agg_grp* + 0.9964 * *year_cov*

The model correctly predicted 76.1% of the observations at a probability level of 0.22. For a probability level of 0.08, the model correctly predicted only 21.4% of the observations (see Table E1 in Appendix E).

Table 4.9 **Evaluations of predictive abilities of models**

Model	% Correctly predicted
Stepwise selection model (with only Year covariate and Aggregate)	76.1
Full model	68.3
Mathematics	67
Aggregate	66.8
Year covariate	56.7
English	65.5
Full model without Year covariate	64.1
African	57.7
Immediately	57.7
Gender	49.6

Note: All models evaluated at probability threshold of 0.22

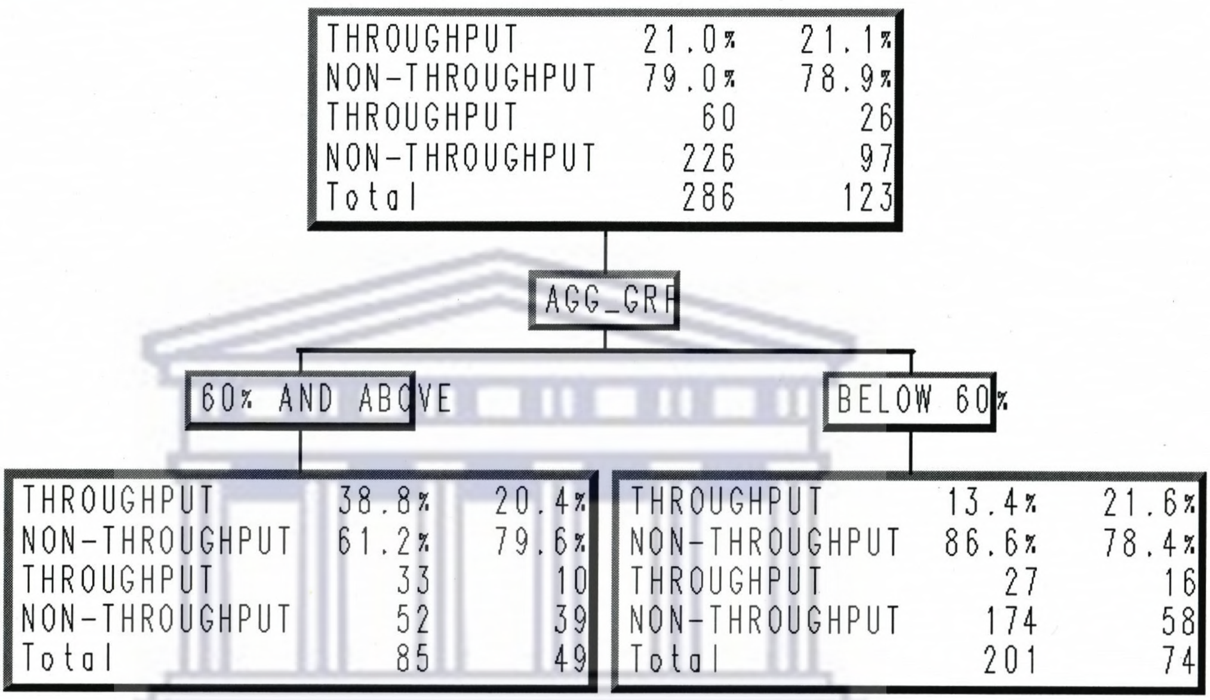
Table 4.9 gives a summary of all the models. The model with only the Year covariate and the Aggregate was the best model to predict throughput. In the following section, the decision tree analysis technique was applied as an alternative way of determining which predictors to include in the model to predict successful throughput.

4.11 Decision Tree Analysis

4.11.1 Aggregate Decision Tree model

The throughput of a student who had an aggregate of 60% and above was 38.8% for the training model and 20.4% for the validation data set, compared to the throughput of a student who had an aggregate of below 60%, which was 13.4% for training data and 21.6% for the validated data set. The validations' modelling throughput for the predictor Aggregate was similar for both validation datasets. The huge difference between the validation and training data sets indicates the instability of the model. The reason for the instability was too few data observations. However, it is interesting to note that the Aggregate predictor was selected although the model is unstable and no statistical interpretation can be inferred from it (see Figure 4.3).

Figure 4.3 **Decision tree aggregate model**



4.12 Conclusion

The decision tree analysis and the stepwise logistic regression both selected the Aggregate predictor as a factor that affected successful throughput. In the following chapter, the findings of Chapter 4 will be discussed and interpreted.

Chapter 5

Discussion and Recommendation

5.1 Discussions of findings

In this thesis, an investigation into how certain factors influence throughput was undertaken. Throughput is the number of students who complete their university studies in the prescribed time. This thesis does not explore the financial and social influences on throughput. The thesis looks at factors like gender, race, home language, Grade 12 aggregate, Grade 12 mathematics, entering UWC immediately after school and the political environment prior and after 1994.

5.1.1 Gender factor

Gender does not play a significant role when investigating throughput. In the study, a significant increase in the number of African female students who enrolled for Statistics was observed.

5.1.2 Race factor

Race is a factor that influences throughput. More non-African students than African students are finishing their studies in the prescribed time.

5.1.3 Home language factor

Home language influences throughput. More non-English students are not finishing their studies in the prescribed time. This shows that if the medium of instruction is different from the students' home language, it can play a role in influencing how long students take to finish their studies.

5.1.4 Grade 12 aggregate factor

The Grade 12 aggregate is the most significant factor influencing throughput. This finding is also confirmed in a study conducted by Lourens and Smit (2003). The aggregate is a factor that should be considered when selecting students, as a higher aggregate relates to better throughput.

5.1. 5 Grade 12 mathematics factor

Mathematics should be made a prerequisite for subjects where calculation and abstract thinking is necessary. The issue is: at what level should students have passed mathematics to be selected for a science subject? This also has enrolment implications in that if the mathematics prerequisite is set too high, the student enrolment in science will drop significantly. If the mathematics prerequisite is set too low, more students with low grades will apply to study in the Science Faculty. Students with low grades in Grade 12 mathematics will take longer to grasp concepts, which will influence the time they take to finish their studies.

5.1.6 Entering UWC immediately after school factor

The impact on throughput of students in the study enrolling at UWC directly after Grade 12 was also investigated. It was found that a short break between school completion and university enrolment does not influence the throughput. In fact, the throughput was better for students who did not enter university immediately after school. The reason could be that older students are more serious about their studies.

5.1.7 Political environment (year covariate) factor

The political environment is one factor which is not often considered in academic studies. It was found that the political change of 1994 did have an impact on throughput. The throughput rate doubled after the 1994 election. However, the intake of students who did at least one semester of third-year Statistics only increased by 17 students after the 1994 election until 2001. The throughput rate doubled most probably because students saw a post-Grade 12 qualification as a means to a brighter future in South Africa.

A logistic regression model was built using the abovementioned factors. It was found that the Grade 12 aggregate and the political environment were the most significant variables to distinguish between students completing their studies in the prescribed time and students taking more that three years. The students in this study were categorized into two groups: those with a Grade 12 aggregate of 60% and above, and those with a Grade 12 aggregate below 60%. The students who had an aggregate of 60% and above had a significantly improved throughput rate compared to those with an aggregate below 60%. The goal is to enroll more students with aggregates higher than 60% so that the throughput rate can be increased.

5.2 Relevance of study

In the study, a model was developed that took into account certain factors that influence university throughput. The model and factors could assist with university policies regarding student selection. Furthermore, minimizing study years would result in students entering the workforce quicker and becoming economically active at an earlier stage. Students could also start sooner with postgraduate studies after successful completion of undergraduate studies.

5.3 Recommendation

As aggregate is an important measure of success at university, it should possibly be retained in the Further Education Training (FET) school system to be implemented in 2006. The new FET system measures a student's performance per subject on a scale of one to five, without an aggregate. Universities should be proactive in formulating new selection criteria systems. A new selection process should be put in place to determine if learners are capable of studying at higher education institutions (HEIs). The entrance requirement for mathematics should also be more strictly enforced to select the best students. It sounds unfair to those students who obtained low grades due to the specific school environment, but higher education institutions cannot repeat the work that the school system should have covered. The responsibility rests on the school teachers and the learners to ensure that learners who want to enroll at university be informed of their choices and how to achieve their goals.

5.4 Limitations of study

The sample was not representative of all the students at UWC as the study was limited to students majoring in Statistics. Historical data was used which was limited to what was on the UWC database. No data were available on students' socio-economic factors, such as financial constraints, mode of transport to university, adequate place to study, and so forth.

5.5 Further research

The study could be replicated to include all students at UWC. Future research could possibly investigate how financial, social and academic factors influence throughput. An interesting question to ask would be: Is South Africa producing enough graduates to meet the labour market demands in terms of specific skills?

In other words, are the targets set by the Department of Education met in terms of graduate output for the new century?

Bibliography

Agar D.L. (1991). *Academic support and academic progress: English-Second-Language speakers in a Faculty of Commerce at a university in South Africa*. Journal of Negro Education Vol. 60(1) : 62 – 77.

Asmal K. (1999). *Call to action: nine priorities of the Minister of Education*. Pretoria, Department of Education.

Badenhorst F.D., Foster D.H. and Lea S.J. (1990). *Factors affecting academic performance in the first-year psychology at the University of Cape Town*. South African Journal of Higher Education (SAJHE) Vol. 4(1) : 39 – 45.

Bargate K. (1999). *Mathematics as an indicator of success in first year Accounting programmes at Technicon Natal*. South African Journal of Higher Education (SAJHE) Vol. 13(1) : 139 – 143.

Bitzer E. and Troskie-De Bruin C. (2004). *The effect of factors related to prior schooling on student persistence in higher education*. South African Journal of Education Vol. 24(2) : 119 – 125.

Botha A.E., McCrindle C.M.E. and Owen J.H. (2003). *Mathematics at matriculation level as an indicator of success or failure in the 1st year of the Veterinary Nursing Diploma at the Faculty of Veterinary Science, University of Pretoria*. Journal of the South African Veterinary Association Vol. 74(4) : 132 – 134.

Cairncross A. (1999). *Progression of students through the Human Ecology Programme (1994 - 1999)*. Department of Human Ecology and Dietetics, UWC. Bellville.

Dawes P., Yeld N. and Smith M.J. (1999). *Access, selection and admission to higher education: maximizing the use of the school-leaving examination*. South African Journal of Higher Education (SAJHE) Vol. 13(3) : 97 – 104.

Department of Education (2001a). *National Plan for Higher Education (NPHE)*. Ministry of Education, Pretoria, February 2001.

Department of Education (2001b). *Funding of Public Higher Education: A New Framework*. Ministry of Education, Pretoria, March 2001.

Fraser W.J. and Killen R. (2003). *Factors influencing academic success or failure of first-year and senior university students: do education students and lectures perceive things differently?* South African Journal of Higher Education (SAJHE) Vol. 23(4) : 254 – 260.

Howie S.J. (2003*). Language and other background factors affecting pupils' performance in mathematics in South Africa.* African Journal of Research in SMT Education Vol. 7 : 1 – 20.

Lourens A. and Smit I.P.J. (2003). *Retention: predicting first-year success*. South African Journal of Higher Education (SAJHE) Vol. 17(2) : 169 – 176.

Mitchell F. (1988). *High school accounting and students' performance in the first level university accounting course: a UK study*. Journal of Accounting Education Vol. 6 : 279 – 291.

Nair P.A.P. (2002). *A theoretical framework for an access programme encompassing further education training: remedy for educational wastage*. South African Journal of Higher Education (SAJHE) Vol. 16 (2) : 94 – 103.

O' Connell B. (2004). *Recapitalisation*. Internal discussion at UWC by Rector, August 2004.

Science Faculty (2004). *UWC Faculty of Science, Undergraduate year book*. UWC.

Skuy M., Zolessi S., Mentis M., Fridjhon P. and Cockcroft K. (1996). *Selection of advantaged and disadvantaged South African students for university admission*. South African Journal of Higher Education (SAJHE) Vol. 10(1) : 110 – 118.

Shochet I. (1985). *University Admission: Can we measure potential?* Paper presented at Academic Support Programme Annual Conference, University of the Witwatersrand.

Stoker D.J., Engelbrecht C.S., Crowther N.A.S., Du Toit S.H.C. and Herbst A. (1986). *Investigation into differential entrance requirement for tertiary educational institutions: an abridged version of Report Ws-31*. Report Number WS-32. Pretoria: Human Science Research Council.

Taylor B. and Harris G. (2002). *The efficiency of South African universities: a study based on the analytical technique.* South African Journal of Higher Education (SAJHE) Vol. 16(2) : 183 – 192.

The Mail and Guardian (1999). *Tertiary state subsidies to change.* 24 November 1999.

Van Rooyen E. (2001). *Die voorspelling van die akademiese prestasie van studente in 'n universiteitsoorbruggings- program.* South African Journal of Higher Education (SAJHE) Vol. 15(1) : 180 – 189.

APPENDIX A

A1 Frequencies of variables

Table A1 GENDER

GENDER	Frequency	Percent	Cumulative Frequency	Cumulative Percent
FEMALE	177	43.28	177	43.28
MALE	232	56.72	409	100.00

Table A2 RACE

RACE	Frequency	Percent	Cumulative Frequency	Cumulative Percent
AFRICAN	230	56.23	230	56.23
COLOURED	156	38.14	386	94.38
INDIAN	22	5.38	408	99.76
WHITE	1	0.24	409	100.00

Table A3 HOMELANG

HOMELANG	Frequency	Percent	Cumulative Frequency	Cumulative Percent
AFRIKAANS	51	12.47	51	12.47
ENG & AFR	35	8.56	86	21.03
ENGLISH	92	22.49	178	43.52
NORTH SOTHO	12	2.93	190	46.45
OTHER	8	1.96	198	48.41
SOUTH SOTHO	17	4.16	215	52.57
SWATI	5	1.22	220	53.79
TSONGA	6	1.47	226	55.26
TSWANA	30	7.33	256	62.59
VENDA	8	1.96	264	64.55
XHOSA	133	32.52	397	97.07
ZULU	12	2.93	409	100.00

Table A4 AFRICAN

AFRICAN	Frequency	Percent	Cumulative Frequency	Cumulative Percent
NON-AFRICAN	179	43.77	179	43.77
AFRICAN	230	56.23	409	100.00

Table A5 ENLGISH

ENGLISH	Frequency	Percent	Cumulative Frequency	Cumulative Percent
NON-ENGLISH	282	68.95	282	68.95
ENGLISH	127	31.05	409	100.00

Table A6

ENDYEAR

ENDYEAR	Frequency	Percent	Cumulative Frequency	Cumulative Percent
1989	2	0.49	2	0.49
1990	8	1.96	10	2.44
1991	2	0.49	12	2.93
1992	9	2.20	21	5.13
1993	12	2.93	33	8.07
1994	17	4.16	50	12.22
1995	19	4.65	69	16.87
1996	40	9.78	109	26.65
1997	48	11.74	157	38.39
1998	43	10.51	200	48.90
1999	48	11.74	248	60.64
2000	46	11.25	294	71.88
2001	32	7.82	326	79.71
2002	39	9.54	365	89.24
2003	28	6.85	393	96.09
2004	16	3.91	409	100.00

Table A7

begyear	Frequency	Percent	Cumulative Frequency	Cumulative Percent
1975	1	0.24	1	0.24
1977	1	0.24	2	0.49
1982	1	0.24	3	0.73
1983	1	0.24	4	0.98
1984	4	0.98	8	1.96
1985	4	0.98	12	2.93
1986	1	0.24	13	3.18
1987	7	1.71	20	4.89
1988	7	1.71	27	6.60
1989	8	1.96	35	8.56
1990	25	6.11	60	14.67
1991	19	4.65	79	19.32
1992	29	7.09	108	26.41
1993	41	10.02	149	36.43
1994	47	11.49	196	47.92
1995	42	10.27	238	58.19
1996	39	9.54	277	67.73
1997	49	11.98	326	79.71
1998	27	6.60	353	86.31
1999	16	3.91	369	90.22
2000	29	7.09	398	97.31
2001	11	2.69	409	100.00

Table A8

year_cov	Frequency	Percent	Cumulative Frequency	Cumulative Percent
PRE-ELECTION YEARS	196	47.92	196	47.92
POST-ELECTION YEARS	213	52.08	409	100.00

Table A9

compl	Frequency	Percent	Cumulative Frequency	Cumulative Percent
3	87	21.27	87	21.27
4	122	29.83	209	51.10
5	100	24.45	309	75.55
6	42	10.27	351	85.82
7	20	4.89	371	90.71
8	14	3.42	385	94.13
9	8	1.96	393	96.09
10	6	1.47	399	97.56
11	3	0.73	402	98.29
12	2	0.49	404	98.78
14	1	0.24	405	99.02
17	2	0.49	407	99.51
18	1	0.24	408	99.76
26	1	0.24	409	100.00

The UNIVARIATE Procedure
Variable: compl

Table A10

Basic Statistical Measures

Location		Variability	
Mean	4.973105	Std Deviation	2.31190
Median	4.000000	Variance	5.34486
Mode	4.000000	Range	23.00000
		Interquartile Range	1.00000

Table A11

through	Frequency	Percent	Cumulative Frequency	Cumulative Percent
NON-THROUGHPUT	323	78.97	323	78.97
THROUGHPUT	86	21.03	409	100.00

Note: Student 2005379 studied three years but did not complete his studies. He repeated his second year and was refused re-entry. Thus 87 - 1 = 86 students who completed their studies.

Table A12

AGG_SYM

AGG_SYM	Frequency	Percent	Cumulative Frequency	Cumulative Percent
A	2	0.50	2	0.50
B	26	6.44	28	6.93
C	106	26.24	134	33.17
D	188	46.53	322	79.70
E	78	19.31	400	99.01
F	4	0.99	404	100.00

Frequency Missing = 5

Table A13

agg_grp	Frequency	Percent	Cumulative Frequency	Cumulative Percent
BELOW 60%	270	66.83	270	66.83
60% AND ABOVE	134	33.17	404	100.00

Frequency Missing = 5

Table A14

MATH_GRD

MATH_GRD	Frequency	Percent	Cumulative Frequency	Cumulative Percent
H	198	48.89	198	48.89
S	207	51.11	405	100.00

Frequency Missing = 4

Table A15

MATH_SYM

MATH_SYM	Frequency	Percent	Cumulative Frequency	Cumulative Percent
A	32	7.90	32	7.90
B	55	13.58	87	21.48
C	72	17.78	159	39.26
D	122	30.12	281	69.38
E	116	28.64	397	98.02
F	8	1.98	405	100.00

Frequency Missing = 4

Table A16

Table of MATH_GRD by MATH_SYM

MATH_GRD(MATH_GRD) MATH_SYM(MATH_SYM)

Frequency / Percent / Row Pct / Col Pct	A	B	C	D	E	F	Total
H	4	17	26	58	92	0	197
	0.99	4.22	6.45	14.39	22.83	0.00	48.88
	2.03	8.63	13.20	29.44	46.70	0.00	
	12.90	30.91	36.62	47.54	79.31	0.00	
S	27	38	45	64	24	8	206
	6.70	9.43	11.17	15.88	5.96	1.99	51.12
	13.11	18.45	21.84	31.07	11.65	3.88	
	87.10	69.09	63.38	52.46	20.69	100.00	
Total	31	55	71	122	116	8	403
	7.69	13.65	17.62	30.27	28.78	1.99	100.00

Frequency Missing = 6

Table A17

common1	Frequency	Percent	Cumulative Frequency	Cumulative Percent
A	4	0.99	4	0.99
B	44	10.92	48	11.91
C	64	15.88	112	27.79
D	103	25.56	215	53.35
E	156	38.71	371	92.06
F	24	5.96	395	98.01
G	8	1.99	403	100.00

Frequency Missing = 6

Table A18

math_grp	Frequency	Percent	Cumulative Frequency	Cumulative Percent
BELOW 60%	291	72.21	291	72.21
60% AND ABOVE	112	27.79	403	100.00

Frequency Missing = 6

Table A19

imed_yrs	Frequency	Percent	Cumulative Frequency	Cumulative Percent
0	238	58.19	238	58.19
1	75	18.34	313	76.53
2	30	7.33	343	83.86
3	30	7.33	373	91.20
4	15	3.67	388	94.87
5	7	1.71	395	96.58
6	5	1.22	400	97.80
7	2	0.49	402	98.29
8	2	0.49	404	98.78
10	2	0.49	406	99.27
14	1	0.24	407	99.51
15	1	0.24	408	99.76
70	1	0.24	409	100.00

Table A20

immediate	Frequency	Percent	Cumulative Frequency	Cumulative Percent
NOT DIRECTLY AFTER SCHOOL	171	41.81	171	41.81
DIRECTLY AFTER SCHOOL	238	58.19	409	100.00

Table A21

subjcode	Frequency	Percent	Cumulative Frequency	Cumulative Percent
MATHEMATICAL STATISTICS	205	50.12	205	50.12
APPLIED STATISTICS	204	49.88	409	100.00

Table A22

pass	Frequency	Percent	Cumulative Frequency	Cumulative Percent
FAIL BOTH SEMESTER	13	3.18	13	3.18
PASS ONLY FIRST SEMESTER	9	2.20	22	5.38
PASS ONLY SECOND SEMESTER	26	6.36	48	11.74
PASS BOTH SEMESTERS	361	88.26	409	100.00

Table A23

major	Frequency	Percent	Cumulative Frequency	Cumulative Percent
N	48	11.74	48	11.74
Y	361	88.26	409	100.00

A2 Throughput associations

Table A24 Association of Throughput by GENDER

Throughput Frequency Percent Row Pct Col Pct	GENDER(GENDER) FEMALE	MALE	Total
NON-THROUGHPUT	147 35.94 45.51 83.05	176 43.03 54.49 75.86	323 78.97
THROUGHPUT	30 7.33 34.88 16.95	56 13.69 65.12 24.14	86 21.03
Total	177 43.28	232 56.72	409 100.00

Statistics for Association of Throughput by GENDER

Statistic	DF	Value	Prob
Chi-Square	1	3.1246	0.0771
Likelihood Ratio Chi-Square	1	3.1735	0.0748
Continuity Adj. Chi-Square	1	2.7067	0.0999
Mantel-Haenszel Chi-Square	1	3.1170	0.0775
Phi Coefficient		0.0874	
Contingency Coefficient		0.0871	
Cramer's V		0.0874	

Fisher's Exact Test

Cell (1,1) Frequency (F)	147
Left-sided Pr <= F	0.9714
Right-sided Pr >= F	0.0492
Table Probability (P)	0.0206
Two-sided Pr <= P	0.0868

Sample Size = 409

Table A25 Association of Throughput by african

Throughput Frequency Percent Row Pct Col Pct	african NON-AFRICAN	AFRICAN	Total
NON-THROUGHPUT	133 32.52 41.18 74.30	190 46.45 58.82 82.61	323 78.97
THROUGHPUT	46 11.25 53.49 25.70	40 9.78 46.51 17.39	86 21.03
Total	179 43.77	230 56.23	409 100.00

Statistics for Association of Throughput by african

Statistic	DF	Value	Prob
Chi-Square	1	4.1831	0.0408
Likelihood Ratio Chi-Square	1	4.1546	0.0415
Continuity Adj. Chi-Square	1	3.6978	0.0545
Mantel-Haenszel Chi-Square	1	4.1728	0.0411
Phi Coefficient		-0.1011	
Contingency Coefficient		0.1006	
Cramer's V		-0.1011	

Fisher's Exact Test

Cell (1,1) Frequency (F)	133
Left-sided Pr <= F	0.0275
Right-sided Pr >= F	0.9847
Table Probability (P)	0.0122
Two-sided Pr <= P	0.0501

Sample Size = 409

Table A26 Association of Throughput by english

Throughput / Frequency / Percent / Row Pct / Col Pct	english: NON-ENGLISH	ENGLISH	Total
NON-THROUGHPUT	232 56.72 71.83 82.27	91 22.25 28.17 71.65	323 78.97
THROUGHPUT	50 12.22 58.14 17.73	36 8.80 41.86 28.35	86 21.03
Total	282 68.95	127 31.05	409 100.00

Statistics for Association of Throughput by english

Statistic	DF	Value	Prob
Chi-Square	1	5.9428	0.0148
Likelihood Ratio Chi-Square	1	5.7270	0.0167
Continuity Adj. Chi-Square	1	5.3207	0.0211
Mantel-Haenszel Chi-Square	1	5.9283	0.0149
Phi Coefficient		0.1205	
Contingency Coefficient		0.1197	
Cramer's V		0.1205	

Fisher's Exact Test

Cell (1,1) Frequency (F)	232
Left-sided Pr <= F	0.9943
Right-sided Pr >= F	0.0115
Table Probability (P)	0.0057
Two-sided Pr <= P	0.0182

Sample Size = 409

Table A27 Association of Throughput by agg_grp

Throughput / Frequency, Percent, Row Pct, Col Pct	agg_grp: BELOW 60 %	60% AND ABOVE	Total
NON-THROUGHPUT	227 56.19 71.38 84.07	91 22.52 28.62 67.91	318 78.71
THROUGHPUT	43 10.64 50.00 15.93	43 10.64 50.00 32.09	86 21.29
Total	270 66.83	134 33.17	404 100.00

Frequency Missing = 5

Statistics for Association of Throughput by agg_grp

Statistic	DF	Value	Prob
Chi-Square	1	13.9637	0.0002
Likelihood Ratio Chi-Square	1	13.3920	0.0003
Continuity Adj. Chi-Square	1	13.0157	0.0003
Mantel-Haenszel Chi-Square	1	13.9292	0.0002
Phi Coefficient		0.1859	
Contingency Coefficient		0.1828	
Cramer's V		0.1859	

Fisher's Exact Test

Cell (1,1) Frequency (F)	227
Left-sided Pr <= F	0.9999
Right-sided Pr >= F	1.998E-04
Table Probability (P)	1.244E-04
Two-sided Pr <= P	2.801E-04

Effective Sample Size = 404
Frequency Missing = 5

Table A28 Association of Throughput by math_grp

Throughput / math_grp Frequency Percent Row Pct Col Pct	BELOW 60 %	60% AND ABOVE	Total
NON-THROUGHPUT	238 59.06 74.84 81.79	80 19.85 25.16 71.43	318 78.91
THROUGHPUT	53 13.15 62.35 18.21	32 7.94 37.65 28.57	85 21.09
Total	291 72.21	112 27.79	403 100.00

Frequency Missing = 6

The FREQ Procedure

Statistics for Association of Throughput by math_grp

Statistic	DF	Value	Prob
Chi-Square	1	5.2138	0.0224
Likelihood Ratio Chi-Square	1	4.9927	0.0255
Continuity Adj. Chi-Square	1	4.6100	0.0318
Mantel-Haenszel Chi-Square	1	5.2008	0.0226
Phi Coefficient		0.1137	
Contingency Coefficient		0.1130	
Cramer's V		0.1137	

Fisher's Exact Test

Cell (1,1) Frequency (F)	238
Left-sided Pr <= F	0.9912
Right-sided Pr >= F	0.0173
Table Probability (P)	0.0085
Two-sided Pr <= P	0.0288

Effective Sample Size = 403
Frequency Missing = 6

Table A29 Association of Throughput by immediate

Throughput / Frequency / Percent / Row Pct / Col Pct	immediate: NOT DIRECTLY AFTER SCHOOL	immediate: DIRECTLY AFTER SCHOOL	Total
NON-THROUGHPUT	129 31.54 39.94 75.44	194 47.43 60.06 81.51	323 78.97
THROUGHPUT	42 10.27 48.84 24.56	44 10.76 51.16 18.49	86 21.03
Total	171 41.81	238 58.19	409 100.00

The FREQ Procedure

Statistics for Association of Throughput by immediate

Statistic	DF	Value	Prob
Chi-Square	1	2.2108	0.1370
Likelihood Ratio Chi-Square	1	2.1916	0.1388
Continuity Adj. Chi-Square	1	1.8601	0.1726
Mantel-Haenszel Chi-Square	1	2.2054	0.1375
Phi Coefficient		-0.0735	
Contingency Coefficient		0.0733	
Cramer's V		-0.0735	

Fisher's Exact Test

Cell (1,1) Frequency (F)	129
Left-sided Pr <= F	0.0867
Right-sided Pr >= F	0.9458
Table Probability (P)	0.0325
Two-sided Pr <= P	0.1420

Sample Size = 409

Table A30 Association of Throughput by year_cov

Throughput (Frequency / Percent / Row Pct / Col Pct)	year_cov: PRE-ELECTION YEARS	POST-ELECTION YEARS	Total
NON-THROUGHPUT	171 41.81 52.94 87.24	152 37.16 47.06 71.36	323 78.97
THROUGHPUT	25 6.11 29.07 12.76	61 14.91 70.93 28.64	86 21.03
Total	196 47.92	213 52.08	409 100.00

The FREQ Procedure

Statistics for Association of Throughput by year_cov

Statistic	DF	Value	Prob
Chi-Square	1	15.5076	<.0001
Likelihood Ratio Chi-Square	1	15.9555	<.0001
Continuity Adj. Chi-Square	1	14.5658	0.0001
Mantel-Haenszel Chi-Square	1	15.4697	<.0001
Phi Coefficient		0.1947	
Contingency Coefficient		0.1911	
Cramer's V		0.1947	

Fisher's Exact Test

Cell (1,1) Frequency (F)	171
Left-sided Pr <= F	1.0000
Right-sided Pr >= F	5.563E-05
Table Probability (P)	3.646E-05
Two-sided Pr <= P	9.032E-05

Sample Size = 409

A3 Predictor associations

Table A31 Association of GENDER by african

GENDER(GENDER) african

Frequency / Percent / Row Pct / Col Pct	NON-AFRICAN	AFRICAN	Total
FEMALE	61	116	177
	14.91	28.36	43.28
	34.46	65.54	
	34.08	50.43	
MALE	118	114	232
	28.85	27.87	56.72
	50.86	49.14	
	65.92	49.57	
Total	179	230	409
	43.77	56.23	100.00

The FREQ Procedure

Statistics for Association of GENDER by african

Statistic	DF	Value	Prob
Chi-Square	1	10.9705	0.0009
Likelihood Ratio Chi-Square	1	11.0696	0.0009
Continuity Adj. Chi-Square	1	10.3143	0.0013
Mantel-Haenszel Chi-Square	1	10.9437	0.0009
Phi Coefficient		-0.1638	
Contingency Coefficient		0.1616	
Cramer's V		-0.1638	

Fisher's Exact Test

Cell (1,1) Frequency (F)	61
Left-sided Pr <= F	6.320E-04
Right-sided Pr >= F	0.9997
Table Probability (P)	3.275E-04
Two-sided Pr <= P	0.0013

Sample Size = 409

Table A32 Association of GENDER by english

GENDER(GENDER) / Frequency / Percent / Row Pct / Col Pct	NON-ENGLISH	ENGLISH	Total
FEMALE	129 31.54 72.88 45.74	48 11.74 27.12 37.80	177 43.28
MALE	153 37.41 65.95 54.26	79 19.32 34.05 62.20	232 56.72
Total	282 68.95	127 31.05	409 100.00

The FREQ Procedure

Statistics for Association of GENDER by english

Statistic	DF	Value	Prob
Chi-Square	1	2.2542	0.1333
Likelihood Ratio Chi-Square	1	2.2704	0.1319
Continuity Adj. Chi-Square	1	1.9420	0.1635
Mantel-Haenszel Chi-Square	1	2.2486	0.1337
Phi Coefficient		0.0742	
Contingency Coefficient		0.0740	
Cramer's V		0.0742	

Fisher's Exact Test

Cell (1,1) Frequency (F)	129
Left-sided Pr <= F	0.9466
Right-sided Pr >= F	0.0814
Table Probability (P)	0.0280
Two-sided Pr <= P	0.1608

Sample Size = 409

Table A33 Association of agg_grp by GENDER

agg_grp Frequency Percent Row Pct Col Pct	GENDER(GENDER) FEMALE	MALE	Total
BELOW 60%	127 31.44 47.04 72.57	143 35.40 52.96 62.45	270 66.83
60% AND ABOVE	48 11.88 35.82 27.43	86 21.29 64.18 37.55	134 33.17
Total	175 43.32	229 56.68	404 100.00

Frequency Missing = 5

The FREQ Procedure

Statistics for Association of agg_grp by GENDER

Statistic	DF	Value	Prob
Chi-Square	1	4.5884	0.0322
Likelihood Ratio Chi-Square	1	4.6346	0.0313
Continuity Adj. Chi-Square	1	4.1430	0.0418
Mantel-Haenszel Chi-Square	1	4.5771	0.0324
Phi Coefficient		0.1066	
Contingency Coefficient		0.1060	
Cramer's V		0.1066	

Fisher's Exact Test

Cell (1,1) Frequency (F)	127
Left-sided Pr <= F	0.9880
Right-sided Pr >= F	0.0205
Table Probability (P)	0.0086
Two-sided Pr <= P	0.0335

Effective Sample Size = 404
Frequency Missing = 5

Table A34 Association of math_grp by GENDER

math_grp GENDER(GENDER)

Frequency Percent Row Pct Col Pct	FEMALE	MALE	Total
BELOW 60%	142 35.24 48.80 80.68	149 36.97 51.20 65.64	291 72.21
60% AND ABOVE	34 8.44 30.36 19.32	78 19.35 69.64 34.36	112 27.79
Total	176 43.67	227 56.33	403 100.00

Frequency Missing = 6

The FREQ Procedure

Statistics for Association of math_grp by GENDER

Statistic	DF	Value	Prob
Chi-Square	1	11.1790	0.0008
Likelihood Ratio Chi-Square	1	11.4573	0.0007
Continuity Adj. Chi-Square	1	10.4420	0.0012
Mantel-Haenszel Chi-Square	1	11.1513	0.0008
Phi Coefficient		0.1666	
Contingency Coefficient		0.1643	
Cramer's V		0.1666	

Fisher's Exact Test

Cell (1,1) Frequency (F)	142
Left-sided Pr <= F	0.9998
Right-sided Pr >= F	5.458E-04
Table Probability (P)	3.103E-04
Two-sided Pr <= P	0.0011

Effective Sample Size = 403
Frequency Missing = 6

Table A35 Association of immediate by GENDER

immediate (Frequency / Percent / Row Pct / Col Pct)	GENDER(GENDER): FEMALE	MALE	Total
NOT DIRECTLY AFTER SCHOOL	59	112	171
	14.43	27.38	41.81
	34.50	65.50	
	33.33	48.28	
DIRECTLY AFTER SCHOOL	118	120	238
	28.85	29.34	58.19
	49.58	50.42	
	66.67	51.72	
Total	177	232	409
	43.28	56.72	100.00

The FREQ Procedure

Statistics for Association of immediate by GENDER

Statistic	DF	Value	Prob
Chi-Square	1	9.2142	0.0024
Likelihood Ratio Chi-Square	1	9.2988	0.0023
Continuity Adj. Chi-Square	1	8.6103	0.0033
Mantel-Haenszel Chi-Square	1	9.1917	0.0024
Phi Coefficient		-0.1501	
Contingency Coefficient		0.1484	
Cramer's V		-0.1501	

Fisher's Exact Test

Cell (1,1) Frequency (F)	59
Left-sided Pr <= F	0.0016
Right-sided Pr >= F	0.9992
Table Probability (P)	7.959E-04
Two-sided Pr <= P	0.0025

Sample Size = 409

Table A36 Association of year_cov by GENDER

year_cov Frequency Percent Row Pct Col Pct	GENDER(GENDER) FEMALE	MALE	Total
PRE-ELECTION YEARS	76 18.58 38.78 42.94	120 29.34 61.22 51.72	196 47.92
POST-ELECTION YEARS	101 24.69 47.42 57.06	112 27.38 52.58 48.28	213 52.08
Total	177 43.28	232 56.72	409 100.00

The FREQ Procedure

Statistics for Association of year_cov by GENDER

Statistic	DF	Value	Prob
Chi-Square	1	3.1057	0.0780
Likelihood Ratio Chi-Square	1	3.1120	0.0777
Continuity Adj. Chi-Square	1	2.7636	0.0964
Mantel-Haenszel Chi-Square	1	3.0981	0.0784
Phi Coefficient		-0.0871	
Contingency Coefficient		0.0868	
Cramer's V		-0.0871	

Fisher's Exact Test

Cell (1,1) Frequency (F)	76
Left-sided Pr <= F	0.0481
Right-sided Pr >= F	0.9688
Table Probability (P)	0.0169
Two-sided Pr <= P	0.0896

Sample Size = 409

Table A37 Association of african by english

african / Frequency, Percent, Row Pct, Col Pct	english: NON-ENGLISH	ENGLISH	Total
NON-AFRICAN	54 13.20 30.17 19.15	125 30.56 69.83 98.43	179 43.77
AFRICAN	228 55.75 99.13 80.85	2 0.49 0.87 1.57	230 56.23
Total	282 68.95	127 31.05	409 100.00

The FREQ Procedure

Statistics for Association of african by english

Statistic	DF	Value	Prob
Chi-Square	1	223.6050	<.0001
Likelihood Ratio Chi-Square	1	264.6022	<.0001
Continuity Adj. Chi-Square	1	220.3955	<.0001
Mantel-Haenszel Chi-Square	1	223.0583	<.0001
Phi Coefficient		-0.7394	
Contingency Coefficient		0.5945	
Cramer's V		-0.7394	

Fisher's Exact Test

Cell (1,1) Frequency (F)	54
Left-sided Pr <= F	1.448E-58
Right-sided Pr >= F	1.0000
Table Probability (P)	1.443E-58
Two-sided Pr <= P	1.448E-58

Sample Size = 409

Table A38 Association of agg_grp by african

agg_grp Frequency Percent Row Pct Col Pct	african NON-AFRICAN	AFRICAN	Total
BELOW 60%	74 18.32 27.41 41.34	196 48.51 72.59 87.11	270 66.83
60% AND ABOVE	105 25.99 78.36 58.66	29 7.18 21.64 12.89	134 33.17
Total	179 44.31	225 55.69	404 100.00

Frequency Missing = 5

The FREQ Procedure

Statistics for Association of agg_grp by african

Statistic	DF	Value	Prob
Chi-Square	1	94.2142	<.0001
Likelihood Ratio Chi-Square	1	97.7024	<.0001
Continuity Adj. Chi-Square	1	92.1607	<.0001
Mantel-Haenszel Chi-Square	1	93.9810	<.0001
Phi Coefficient		-0.4829	
Contingency Coefficient		0.4349	
Cramer's V		-0.4829	

Fisher's Exact Test

Cell (1,1) Frequency (F)	74
Left-sided Pr <= F	7.696E-23
Right-sided Pr >= F	1.0000
Table Probability (P)	6.911E-23
Two-sided Pr <= P	1.209E-22

Effective Sample Size = 404
Frequency Missing = 5

Table A39 Association of math_grp by african

math_grp / Frequency, Percent, Row Pct, Col Pct	NON-AFRICAN	AFRICAN	Total
BELOW 60%	97 24.07 33.33 55.11	194 48.14 66.67 85.46	291 72.21
60% AND ABOVE	79 19.60 70.54 44.89	33 8.19 29.46 14.54	112 27.79
Total	176 43.67	227 56.33	403 100.00

Frequency Missing = 6

The FREQ Procedure

Statistics for Association of math_grp by african

Statistic	DF	Value	Prob
Chi-Square	1	45.5008	<.0001
Likelihood Ratio Chi-Square	1	45.9525	<.0001
Continuity Adj. Chi-Square	1	44.0010	<.0001
Mantel-Haenszel Chi-Square	1	45.3879	<.0001
Phi Coefficient		-0.3360	
Contingency Coefficient		0.3185	
Cramer's V		-0.3360	

Fisher's Exact Test

Cell (1,1) Frequency (F)	97
Left-sided Pr <= F	1.344E-11
Right-sided Pr >= F	1.0000
Table Probability (P)	1.072E-11
Two-sided Pr <= P	1.734E-11

Effective Sample Size = 403
Frequency Missing = 6

Table A40 Association of immediate by african

immediate by african

Frequency Percent Row Pct Col Pct	NON-AFRICAN	AFRICAN	Total
NOT DIRECTLY AFTER SCHOOL	57 13.94 33.33 31.84	114 27.87 66.67 49.57	171 41.81
DIRECTLY AFTER SCHOOL	122 29.83 51.26 68.16	116 28.36 48.74 50.43	238 58.19
Total	179 43.77	230 56.23	409 100.00

The FREQ Procedure

Statistics for Association of immediate by african

Statistic	DF	Value	Prob
Chi-Square	1	12.9939	0.0003
Likelihood Ratio Chi-Square	1	13.1438	0.0003
Continuity Adj. Chi-Square	1	12.2757	0.0005
Mantel-Haenszel Chi-Square	1	12.9621	0.0003
Phi Coefficient		-0.1782	
Contingency Coefficient		0.1755	
Cramer's V		-0.1782	

Fisher's Exact Test

Cell (1,1) Frequency (F)	57
Left-sided Pr <= F	2.147E-04
Right-sided Pr >= F	0.9999
Table Probability (P)	1.176E-04
Two-sided Pr <= P	3.927E-04

Sample Size = 409

Table A41 Association of year_cov by african

year_cov / Frequency / Percent / Row Pct / Col Pct	african: NON-AFRICAN	AFRICAN	Total
PRE-ELECTION YEARS	94 22.98 47.96 52.51	102 24.94 52.04 44.35	196 47.92
POST-ELECTION YEARS	85 20.78 39.91 47.49	128 31.30 60.09 55.65	213 52.08
Total	179 43.77	230 56.23	409 100.00

The FREQ Procedure

Statistics for Association of year_cov by african

Statistic	DF	Value	Prob
Chi-Square	1	2.6897	0.1010
Likelihood Ratio Chi-Square	1	2.6913	0.1009
Continuity Adj. Chi-Square	1	2.3724	0.1235
Mantel-Haenszel Chi-Square	1	2.6831	0.1014
Phi Coefficient		0.0811	
Contingency Coefficient		0.0808	
Cramer's V		0.0811	

Fisher's Exact Test

Cell (1,1) Frequency (F)	94
Left-sided Pr <= F	0.9591
Right-sided Pr >= F	0.0617
Table Probability (P)	0.0208
Two-sided Pr <= P	0.1109

Sample Size = 409

Table A42 Association of agg_grp by english

agg_grp Frequency Percent Row Pct Col Pct	english NON-ENGLISH	ENGLISH	Total
BELOW 60%	213 52.72 78.89 76.90	57 14.11 21.11 44.88	270 66.83
60% AND ABOVE	64 15.84 47.76 23.10	70 17.33 52.24 55.12	134 33.17
Total	277 68.56	127 31.44	404 100.00

Frequency Missing = 5

The FREQ Procedure

Statistics for Association of agg_grp by english

Statistic	DF	Value	Prob
Chi-Square	1	40.2588	<.0001
Likelihood Ratio Chi-Square	1	39.1898	<.0001
Continuity Adj. Chi-Square	1	38.8275	<.0001
Mantel-Haenszel Chi-Square	1	40.1591	<.0001
Phi Coefficient		0.3157	
Contingency Coefficient		0.3010	
Cramer's V		0.3157	

Fisher's Exact Test

Cell (1,1) Frequency (F)	213
Left-sided Pr <= F	1.0000
Right-sided Pr >= F	3.876E-10
Table Probability (P)	2.960E-10
Two-sided Pr <= P	4.708E-10

Effective Sample Size = 404
Frequency Missing = 5

Table A43 Association of math_grp by english

math_grp / english Frequency Percent Row Pct Col Pct	NON-ENGLISH	ENGLISH	Total
BELOW 60%	212 52.61 72.85 76.53	79 19.60 27.15 62.70	291 72.21
60% AND ABOVE	65 16.13 58.04 23.47	47 11.66 41.96 37.30	112 27.79
Total	277 68.73	126 31.27	403 100.00

Frequency Missing = 6

The FREQ Procedure

Statistics for Association of math_grp by english

Statistic	DF	Value	Prob
Chi-Square	1	8.2615	0.0040
Likelihood Ratio Chi-Square	1	8.0258	0.0046
Continuity Adj. Chi-Square	1	7.5864	0.0059
Mantel-Haenszel Chi-Square	1	8.2410	0.0041
Phi Coefficient		0.1432	
Contingency Coefficient		0.1417	
Cramer's V		0.1432	

Fisher's Exact Test

Cell (1,1) Frequency (F)	212
Left-sided Pr <= F	0.9984
Right-sided Pr >= F	0.0033
Table Probability (P)	0.0017
Two-sided Pr <= P	0.0056

Effective Sample Size = 403
Frequency Missing = 6

Table A44 Association of immediate by english

immediate (Frequency / Percent / Row Pct / Col Pct)	english: NON-ENGLISH	ENGLISH	Total
NOT DIRECTLY AFTER SCHOOL	127 31.05 74.27 45.04	44 10.76 25.73 34.65	171 41.81
DIRECTLY AFTER SCHOOL	155 37.90 65.13 54.96	83 20.29 34.87 65.35	238 58.19
Total	282 68.95	127 31.05	409 100.00

The FREQ Procedure

Statistics for Association of immediate by english

Statistic	DF	Value	Prob
Chi-Square	1	3.8852	0.0487
Likelihood Ratio Chi-Square	1	3.9316	0.0474
Continuity Adj. Chi-Square	1	3.4699	0.0625
Mantel-Haenszel Chi-Square	1	3.8757	0.0490
Phi Coefficient		0.0975	
Contingency Coefficient		0.0970	
Cramer's V		0.0975	

Fisher's Exact Test

Cell (1,1) Frequency (F)	127
Left-sided Pr <= F	0.9817
Right-sided Pr >= F	0.0307
Table Probability (P)	0.0124
Two-sided Pr <= P	0.0518

Sample Size = 409

Table A45 Association of year_cov by english

year_cov by english

Frequency / Percent / Row Pct / Col Pct	NON-ENGLISH	ENGLISH	Total
PRE-ELECTION YEARS	132 32.27 67.35 46.81	64 15.65 32.65 50.39	196 47.92
POST-ELECTION YEARS	150 36.67 70.42 53.19	63 15.40 29.58 49.61	213 52.08
Total	282 68.95	127 31.05	409 100.00

The FREQ Procedure

Statistics for Association of year_cov by english

Statistic	DF	Value	Prob
Chi-Square	1	0.4510	0.5019
Likelihood Ratio Chi-Square	1	0.4508	0.5020
Continuity Adj. Chi-Square	1	0.3188	0.5723
Mantel-Haenszel Chi-Square	1	0.4499	0.5024
Phi Coefficient		-0.0332	
Contingency Coefficient		0.0332	
Cramer's V		-0.0332	

Fisher's Exact Test

Cell (1,1) Frequency (F)	132
Left-sided Pr <= F	0.2861
Right-sided Pr >= F	0.7819
Table Probability (P)	0.0680
Two-sided Pr <= P	0.5223

Sample Size = 409

Table A46 Association of agg_grp by math_grp

agg_grp (Frequency / Percent / Row Pct / Col Pct)	math_grp: BELOW 60 %	60% AND ABOVE	Total
BELOW 60%	222 55.78 83.77 77.62	43 10.80 16.23 38.39	265 66.58
60% AND ABOVE	64 16.08 48.12 22.38	69 17.34 51.88 61.61	133 33.42
Total	286 71.86	112 28.14	398 100.00

Frequency Missing = 11

The FREQ Procedure

Statistics for Association of agg_grp by math_grp

Statistic	DF	Value	Prob
Chi-Square	1	55.6667	<.0001
Likelihood Ratio Chi-Square	1	53.8508	<.0001
Continuity Adj. Chi-Square	1	53.9176	<.0001
Mantel-Haenszel Chi-Square	1	55.5269	<.0001
Phi Coefficient		0.3740	
Contingency Coefficient		0.3503	
Cramer's V		0.3740	

Fisher's Exact Test

Cell (1,1) Frequency (F)	222
Left-sided Pr <= F	1.0000
Right-sided Pr >= F	2.530E-13
Table Probability (P)	2.089E-13
Two-sided Pr <= P	3.103E-13

Effective Sample Size = 398
Frequency Missing = 11

Table A47 Association of agg_grp by immediate

agg_grp / Frequency, Percent, Row Pct, Col Pct	immediate: NOT DIRECTLY AFTER SCHOOL	immediate: DIRECTLY AFTER SCHOOL	Total
BELOW 60%	116 28.71 42.96 68.24	154 38.12 57.04 65.81	270 66.83
60% AND ABOVE	54 13.37 40.30 31.76	80 19.80 59.70 34.19	134 33.17
Total	170 42.08	234 57.92	404 100.00

Frequency Missing = 5

The FREQ Procedure

Statistics for Association of agg_grp by immediate

Statistic	DF	Value	Prob
Chi-Square	1	0.2609	0.6095
Likelihood Ratio Chi-Square	1	0.2614	0.6092
Continuity Adj. Chi-Square	1	0.1630	0.6864
Mantel-Haenszel Chi-Square	1	0.2602	0.6100
Phi Coefficient		0.0254	
Contingency Coefficient		0.0254	
Cramer's V		0.0254	

Fisher's Exact Test

Cell (1,1) Frequency (F)	116
Left-sided Pr <= F	0.7312
Right-sided Pr >= F	0.3438
Table Probability (P)	0.0750
Two-sided Pr <= P	0.6688

Effective Sample Size = 404
Frequency Missing = 5

Table A48 Association of agg_grp by year_cov

agg_grp Frequency Percent Row Pct Col Pct	year_cov PRE-ELECTION YEARS	POST-ELECTION YEARS	Total
BELOW 60%	131 32.43 48.52 68.59	139 34.41 51.48 65.26	270 66.83
60% AND ABOVE	60 14.85 44.78 31.41	74 18.32 55.22 34.74	134 33.17
Total	191 47.28	213 52.72	404 100.00

Frequency Missing = 5

The FREQ Procedure

Statistics for Association of agg_grp by year_cov

Statistic	DF	Value	Prob
Chi-Square	1	0.5032	0.4781
Likelihood Ratio Chi-Square	1	0.5038	0.4778
Continuity Adj. Chi-Square	1	0.3643	0.5462
Mantel-Haenszel Chi-Square	1	0.5020	0.4786
Phi Coefficient		0.0353	
Contingency Coefficient		0.0353	
Cramer's V		0.0353	

Fisher's Exact Test

Cell (1,1) Frequency (F)	131
Left-sided Pr <= F	0.7924
Right-sided Pr >= F	0.2733
Table Probability (P)	0.0657
Two-sided Pr <= P	0.5258

Effective Sample Size = 404
Frequency Missing = 5

Table A49 Association of math_grp by immediate

math_grp / Frequency, Percent, Row Pct, Col Pct	immediate: NOT DIRECTLY AFTER SCHOOL	DIRECTLY AFTER SCHOOL	Total
BELOW 60%	127 31.51 43.64 75.60	164 40.69 56.36 69.79	291 72.21
60% AND ABOVE	41 10.17 36.61 24.40	71 17.62 63.39 30.21	112 27.79
Total	168 41.69	235 58.31	403 100.00

Frequency Missing = 6

The FREQ Procedure

Statistics for Association of math_grp by immediate

Statistic	DF	Value	Prob
Chi-Square	1	1.6467	0.1994
Likelihood Ratio Chi-Square	1	1.6611	0.1975
Continuity Adj. Chi-Square	1	1.3700	0.2418
Mantel-Haenszel Chi-Square	1	1.6427	0.2000
Phi Coefficient		0.0639	
Contingency Coefficient		0.0638	
Cramer's V		0.0639	

Fisher's Exact Test

Cell (1,1) Frequency (F)	127
Left-sided Pr <= F	0.9191
Right-sided Pr >= F	0.1207
Table Probability (P)	0.0398
Two-sided Pr <= P	0.2160

Effective Sample Size = 403
Frequency Missing = 6

Table A50 Association of math_grp by year_cov

math_grp (Frequency / Percent / Row Pct / Col Pct)	year_cov: PRE-ELECTION YEARS	POST-ELECTION YEARS	Total
BELOW 60%	130 32.26 44.67 68.42	161 39.95 55.33 75.59	291 72.21
60% AND ABOVE	60 14.89 53.57 31.58	52 12.90 46.43 24.41	112 27.79
Total	190 47.15	213 52.85	403 100.00

Frequency Missing = 6

The FREQ Procedure

Statistics for Association of math_grp by year_cov

Statistic	DF	Value	Prob
Chi-Square	1	2.5695	0.1089
Likelihood Ratio Chi-Square	1	2.5672	0.1091
Continuity Adj. Chi-Square	1	2.2249	0.1358
Mantel-Haenszel Chi-Square	1	2.5632	0.1094
Phi Coefficient		-0.0799	
Contingency Coefficient		0.0796	
Cramer's V		-0.0799	

Fisher's Exact Test

Cell (1,1) Frequency (F)	130
Left-sided Pr <= F	0.0680
Right-sided Pr >= F	0.9567
Table Probability (P)	0.0247
Two-sided Pr <= P	0.1195

Effective Sample Size = 403
Frequency Missing = 6

Table A51 Association of immediate by year_cov

immediate / Frequency / Percent / Row Pct / Col Pct	year_cov: PRE-ELECTION YEARS	POST-ELECTION YEARS	Total
NOT DIRECTLY AFTER SCHOOL	92 22.49 53.80 46.94	79 19.32 46.20 37.09	171 41.81
DIRECTLY AFTER SCHOOL	104 25.43 43.70 53.06	134 32.76 56.30 62.91	238 58.19
Total	196 47.92	213 52.08	409 100.00

The FREQ Procedure

Statistics for Association of immediate by year_cov

Statistic	DF	Value	Prob
Chi-Square	1	4.0702	0.0436
Likelihood Ratio Chi-Square	1	4.0740	0.0435
Continuity Adj. Chi-Square	1	3.6755	0.0552
Mantel-Haenszel Chi-Square	1	4.0603	0.0439
Phi Coefficient		0.0998	
Contingency Coefficient		0.0993	
Cramer's V		0.0998	

Fisher's Exact Test

Cell (1,1) Frequency (F)	92
Left-sided Pr <= F	0.9829
Right-sided Pr >= F	0.0276
Table Probability (P)	0.0105
Two-sided Pr <= P	0.0455

Sample Size = 409

Appendix B

B1 Gender logistic regression model

Table B1

GENDER

The LOGISTIC Procedure

Model Information

Data Set	COM.ALL
Response Variable	through
Number of Response Levels	2
Number of Observations	409
Model	binary logit
Optimization Technique	Fisher's scoring

Response Profile

Ordered Value	through	Total Frequency
1	THROUGHPUT	86
2	NON-THROUGHPUT	323

Probability modeled is through='THROUGHPUT'.

Class Level Information

Class	Value	Design Variables 1
GENDER	FEMALE	1
	MALE	-1

Model Convergence Status

Convergence criterion (GCONV=1E-8) satisfied.

Model Fit Statistics

Criterion	Intercept Only	Intercept and Covariates
AIC	422.708	421.534
SC	426.722	429.562
-2 Log L	420.708	417.534

The LOGISTIC Procedure

Testing Global Null Hypothesis: BETA=0

Test	Chi-Square	DF	Pr > ChiSq
Likelihood Ratio	3.1735	1	0.0748
Score	3.1246	1	0.0771
Wald	3.0960	1	0.0785

Type III Analysis of Effects

Effect	DF	Wald Chi-Square	Pr > ChiSq
GENDER	1	3.0960	0.0785

Analysis of Maximum Likelihood Estimates

Parameter		DF	Estimate	Standard Error	Wald Chi-Square	Pr > ChiSq
Intercept		1	-1.3671	0.1262	117.4143	<.0001
GENDER	FEMALE	1	-0.2220	0.1262	3.0960	0.0785

Odds Ratio Estimates

Effect	Point Estimate	95% Wald Confidence Limits	
GENDER FEMALE vs MALE	0.641	0.391	1.052

Association of Predicted Probabilities and Observed Responses

Percent Concordant	29.6	Somers' D	0.106
Percent Discordant	19.0	Gamma	0.218
Percent Tied	51.4	Tau-a	0.035
Pairs	27778	c	0.553

Wald Confidence Interval for Adjusted Odds Ratios

Effect	Unit	Estimate	95% Confidence Limits	
GENDER FEMALE vs MALE	1.0000	0.641	0.391	1.052

The LOGISTIC Procedure

Classification Table

	Correct		Incorrect		Percentages				
Prob Level	Event	Non-Event	Event	Non-Event	Correct	Sensi-tivity	Speci-ficity	False POS	False NEG
0.160	**86**	**0**	**323**	**0**	**21.0**	**100.0**	**0.0**	**79.0**	**.**
0.180	56	147	176	30	49.6	65.1	45.5	75.9	16.9
0.200	56	147	176	30	49.6	65.1	45.5	75.9	16.9
0.220	**56**	**147**	**176**	**30**	**49.6**	**65.1**	**45.5**	**75.9**	**16.9**
0.240	0	147	176	86	35.9	0.0	45.5	100.0	36.9
0.260	0	323	0	86	79.0	0.0	100.0	.	21.0

B2 African logistic regression model

Table B2

AFRICAN

The LOGISTIC Procedure

Model Information

Data Set	COM.ALL
Response Variable	through
Number of Response Levels	2
Number of Observations	409
Model	binary logit
Optimization Technique	Fisher's scoring

Response Profile

Ordered Value	through	Total Frequency
1	THROUGHPUT	86
2	NON-THROUGHPUT	323

Probability modeled is through='THROUGHPUT'.

Model Convergence Status

Convergence criterion (GCONV=1E-8) satisfied.

Model Fit Statistics

Criterion	Intercept Only	Intercept and Covariates
AIC	422.708	420.553
SC	426.722	428.581
-2 Log L	420.708	416.553

Testing Global Null Hypothesis: BETA=0

Test	Chi-Square	DF	Pr > ChiSq
Likelihood Ratio	4.1546	1	0.0415
Score	4.1831	1	0.0408
Wald	4.1409	1	0.0419

The LOGISTIC Procedure

Analysis of Maximum Likelihood Estimates

Parameter	DF	Estimate	Standard Error	Wald Chi-Square	Pr > ChiSq
Intercept	1	-1.0616	0.1710	38.5221	<.0001
african	1	-0.4964	0.2440	4.1409	0.0419

Odds Ratio Estimates

Effect	Point Estimate	95% Wald Confidence Limits	
african	0.609	0.377	0.982

Association of Predicted Probabilities and Observed Responses

Percent Concordant	31.5	Somers' D	0.123
Percent Discordant	19.2	Gamma	0.243
Percent Tied	49.4	Tau-a	0.041
Pairs	27778	c	0.562

Wald Confidence Interval for Adjusted Odds Ratios

Effect	Unit	Estimate	95% Confidence Limits	
african	1.0000	0.609	0.377	0.982

Classification Table

Prob Level	Correct Event	Correct Non-Event	Incorrect Event	Incorrect Non-Event	Percentages Correct	Sensi-tivity	Speci-ficity	False POS	False NEG
0.160	86	0	323	0	21.0	100.0	0.0	79.0	.
0.180	46	190	133	40	57.7	53.5	58.8	74.3	17.4
0.200	46	190	133	40	57.7	53.5	58.8	74.3	17.4
0.220	46	190	133	40	57.7	53.5	58.8	74.3	17.4
0.240	46	190	133	40	57.7	53.5	58.8	74.3	17.4
0.260	0	323	0	86	79.0	0.0	100.0	.	21.0

B3 English logistic regression model

Table B3

ENGLISH
The LOGISTIC Procedure

Model Information

Data Set	COM.ALL
Response Variable	through
Number of Response Levels	2
Number of Observations	409
Model	binary logit
Optimization Technique	Fisher's scoring

Response Profile

Ordered Value	through	Total Frequency
1	THROUGHPUT	86
2	NON-THROUGHPUT	323

Probability modeled is through='THROUGHPUT'.

Model Convergence Status

Convergence criterion (GCONV=1E-8) satisfied.

Model Fit Statistics

Criterion	Intercept Only	Intercept and Covariates
AIC	422.708	418.981
SC	426.722	427.008
-2 Log L	420.708	414.981

Testing Global Null Hypothesis: BETA=0

Test	Chi-Square	DF	Pr > ChiSq
Likelihood Ratio	5.7270	1	0.0167
Score	5.9428	1	0.0148
Wald	5.8484	1	0.0156

The LOGISTIC Procedure

Analysis of Maximum Likelihood Estimates

Parameter	DF	Estimate	Standard Error	Wald Chi-Square	Pr > ChiSq
Intercept	1	-1.5347	0.1559	96.8867	<.0001
english	1	0.6074	0.2512	5.8484	0.0156

Odds Ratio Estimates

Effect	Point Estimate	95% Wald Confidence Limits	
english	1.836	1.122	3.003

Association of Predicted Probabilities and Observed Responses

Percent Concordant	30.1	Somers' D	0.137
Percent Discordant	16.4	Gamma	0.295
Percent Tied	53.6	Tau-a	0.046
Pairs	27778	c	0.568

Wald Confidence Interval for Adjusted Odds Ratios

Effect	Unit	Estimate	95% Confidence Limits	
english	1.0000	1.836	1.122	3.003

Classification Table

Prob Level	Correct Event	Correct Non-Event	Incorrect Event	Incorrect Non-Event	Percentages Correct	Sensi-tivity	Speci-ficity	False POS	False NEG
0.160	86	0	323	0	21.0	100.0	0.0	79.0	.
0.180	36	232	91	50	65.5	41.9	71.8	71.7	17.7
0.200	36	232	91	50	65.5	41.9	71.8	71.7	17.7
0.220	36	232	91	50	65.5	41.9	71.8	71.7	17.7
0.240	36	232	91	50	65.5	41.9	71.8	71.7	17.7
0.260	36	232	91	50	65.5	41.9	71.8	71.7	17.7
0.280	0	232	91	86	56.7	0.0	71.8	100.0	27.0
0.300	0	323	0	86	79.0	0.0	100.0	.	21.0

B4 Aggregate logistic regression model

Table B4

AGGREAGTE

The LOGISTIC Procedure

Model Information

Data Set	COM.ALL
Response Variable	through
Number of Response Levels	2
Number of Observations	404
Model	binary logit
Optimization Technique	Fisher's scoring

Response Profile

Ordered Value	through	Total Frequency
1	THROUGHPUT	86
2	NON-THROUGHPUT	318

Probability modeled is through='THROUGHPUT'.

NOTE: 5 observations were deleted due to missing values for the response or explanatory variables

Model Convergence Status

Convergence criterion (GCONV=1E-8) satisfied.

Model Fit Statistics

Criterion	Intercept Only	Intercept and Covariates
AIC	420.331	408.939
SC	424.332	416.942
-2 Log L	418.331	404.939

Testing Global Null Hypothesis: BETA=0

Test	Chi-Square	DF	Pr > ChiSq
Likelihood Ratio	13.3920	1	0.0003
Score	13.9637	1	0.0002
Wald	13.4973	1	0.0002

The LOGISTIC Procedure

Analysis of Maximum Likelihood Estimates

Parameter	DF	Estimate	Standard Error	Wald Chi-Square	Pr > ChiSq
Intercept	1	-1.6637	0.1663	100.0706	<.0001
agg_grp	1	0.9141	0.2488	13.4973	0.0002

Odds Ratio Estimates

Effect	Point Estimate	95% Wald Confidence Limits	
agg_grp	2.495	1.532	4.062

Association of Predicted Probabilities and Observed Responses

Percent Concordant	35.7	Somers' D	0.214
Percent Discordant	14.3	Gamma	0.428
Percent Tied	50.0	Tau-a	0.072
Pairs	27348	c	0.607

Wald Confidence Interval for Adjusted Odds Ratios

Effect	Unit	Estimate	95% Confidence Limits	
agg_grp	1.0000	2.495	1.532	4.062

Classification Table

Prob Level	Correct Event	Correct Non-Event	Incorrect Event	Incorrect Non-Event	Percentages Correct	Sensi-tivity	Speci-ficity	False POS	False NEG
0.140	86	0	318	0	21.3	100.0	0.0	78.7	.
0.160	43	227	91	43	66.8	50.0	71.4	67.9	15.9
0.180	43	227	91	43	66.8	50.0	71.4	67.9	15.9
0.200	43	227	91	43	66.8	50.0	71.4	67.9	15.9
0.220	43	227	91	43	66.8	50.0	71.4	67.9	15.9
0.240	43	227	91	43	66.8	50.0	71.4	67.9	15.9
0.260	43	227	91	43	66.8	50.0	71.4	67.9	15.9
0.280	43	227	91	43	66.8	50.0	71.4	67.9	15.9
0.300	43	227	91	43	66.8	50.0	71.4	67.9	15.9
0.320	0	227	91	86	56.2	0.0	71.4	100.0	27.5
0.340	0	318	0	86	78.7	0.0	100.0	.	21.3

B5 Mathematics logistic regression model

Table B5

MATHEMATICS
The LOGISTIC Procedure

Model Information

Data Set	COM.ALL
Response Variable	through
Number of Response Levels	2
Number of Observations	403
Model	binary logit
Optimization Technique	Fisher's scoring

Response Profile

Ordered Value	through	Total Frequency
1	THROUGHPUT	85
2	NON-THROUGHPUT	318

Probability modeled is through='THROUGHPUT'.

NOTE: 6 observations were deleted due to missing values for the response or explanatory variables

Model Convergence Status

Convergence criterion (GCONV=1E-8) satisfied.

Model Fit Statistics

Criterion	Intercept Only	Intercept and Covariates
AIC	417.227	414.235
SC	421.226	422.233
-2 Log L	415.227	410.235

Testing Global Null Hypothesis: BETA=0

Test	Chi-Square	DF	Pr > ChiSq
Likelihood Ratio	4.9927	1	0.0255
Score	5.2138	1	0.0224
Wald	5.1337	1	0.0235

The LOGISTIC Procedure

Analysis of Maximum Likelihood Estimates

Parameter	DF	Estimate	Standard Error	Wald Chi-Square	Pr > ChiSq
Intercept	1	-1.5020	0.1519	97.7884	<.0001
math_grp	1	0.5857	0.2585	5.1337	0.0235

Odds Ratio Estimates

Effect	Point Estimate	95% Wald Confidence Limits	
math_grp	1.796	1.082	2.981

Association of Predicted Probabilities and Observed Responses

Percent Concordant	28.2	Somers' D	0.125
Percent Discordant	15.7	Gamma	0.285
Percent Tied	56.1	Tau-a	0.042
Pairs	27030	c	0.562

Wald Confidence Interval for Adjusted Odds Ratios

Effect	Unit	Estimate	95% Confidence Limits	
math_grp	1.0000	1.796	1.082	2.981

Classification Table

	Correct		Incorrect		Percentages				
Prob Level	Event	Non-Event	Event	Non-Event	Correct	Sensi-tivity	Speci-ficity	False POS	False NEG
0.160	85	0	318	0	21.1	100.0	0.0	78.9	.
0.180	32	0	318	53	7.9	37.6	0.0	90.9	100.0
0.200	32	238	80	53	67.0	37.6	74.8	71.4	18.2
0.220	32	238	80	53	67.0	37.6	74.8	71.4	18.2
0.240	32	238	80	53	67.0	37.6	74.8	71.4	18.2
0.260	32	238	80	53	67.0	37.6	74.8	71.4	18.2
0.280	0	238	80	85	59.1	0.0	74.8	100.0	26.3
0.300	0	318	0	85	78.9	0.0	100.0	.	21.1

B6 Immediately logistic regression model

Table B6

IMMEDIATELY

The LOGISTIC Procedure

Model Information

Data Set	COM.ALL
Response Variable	through
Number of Response Levels	2
Number of Observations	409
Model	binary logit
Optimization Technique	Fisher's scoring

Response Profile

Ordered Value	through	Total Frequency
1	THROUGHPUT	86
2	NON-THROUGHPUT	323

Probability modeled is through='THROUGHPUT'.

Model Convergence Status

Convergence criterion (GCONV=1E-8) satisfied.

Model Fit Statistics

Criterion	Intercept Only	Intercept and Covariates
AIC	422.708	422.516
SC	426.722	430.544
-2 Log L	420.708	418.516

Testing Global Null Hypothesis: BETA=0

Test	Chi-Square	DF	Pr > ChiSq
Likelihood Ratio	2.1916	1	0.1388
Score	2.2108	1	0.1370
Wald	2.1990	1	0.1381

The LOGISTIC Procedure

Analysis of Maximum Likelihood Estimates

Parameter	DF	Estimate	Standard Error	Wald Chi-Square	Pr > ChiSq
Intercept	1	-1.1221	0.1777	39.8951	<.0001
immediate	1	-0.3615	0.2438	2.1990	0.1381

Odds Ratio Estimates

Effect	Point Estimate	95% Wald Confidence Limits	
immediate	0.697	0.432	1.123

Association of Predicted Probabilities and Observed Responses

Percent Concordant	29.3	Somers' D	0.089
Percent Discordant	20.4	Gamma	0.179
Percent Tied	50.2	Tau-a	0.030
Pairs	27778	c	0.544

Wald Confidence Interval for Adjusted Odds Ratios

Effect	Unit	Estimate	95% Confidence Limits	
immediate	1.0000	0.697	0.432	1.123

Classification Table

Prob Level	Correct Event	Correct Non-Event	Incorrect Event	Incorrect Non-Event	Percentages Correct	Sensi-tivity	Speci-ficity	False POS	False NEG
0.180	86	0	323	0	21.0	100.0	0.0	79.0	.
0.200	42	194	129	44	57.7	48.8	60.1	75.4	18.5
0.220	42	194	129	44	57.7	48.8	60.1	75.4	18.5
0.240	42	194	129	44	57.7	48.8	60.1	75.4	18.5
0.260	0	323	0	86	79.0	0.0	100.0	.	21.0

B7 Year covariate logistic regression model

Table B7

YEAR COVARIATE
The LOGISTIC Procedure

Model Information

Data Set	COM.ALL
Response Variable	through
Number of Response Levels	2
Number of Observations	409
Model	binary logit
Optimization Technique	Fisher's scoring

Response Profile

Ordered Value	through	Total Frequency
1	THROUGHPUT	86
2	NON-THROUGHPUT	323

Probability modeled is through='THROUGHPUT'.

Model Convergence Status

Convergence criterion (GCONV=1E-8) satisfied.

Model Fit Statistics

Criterion	Intercept Only	Intercept and Covariates
AIC	422.708	408.752
SC	426.722	416.780
-2 Log L	420.708	404.752

Testing Global Null Hypothesis: BETA=0

Test	Chi-Square	DF	Pr > ChiSq
Likelihood Ratio	15.9555	1	<.0001
Score	15.5076	1	<.0001
Wald	14.8161	1	0.0001

The LOGISTIC Procedure

Analysis of Maximum Likelihood Estimates

Parameter	DF	Estimate	Standard Error	Wald Chi-Square	Pr > ChiSq
Intercept	1	-1.9228	0.2141	80.6384	<.0001
year_cov	1	1.0098	0.2623	14.8161	0.0001

Odds Ratio Estimates

Effect	Point Estimate	95% Wald Confidence Limits	
year_cov	2.745	1.641	4.590

Association of Predicted Probabilities and Observed Responses

Percent Concordant	37.6	Somers' D	0.239
Percent Discordant	13.7	Gamma	0.466
Percent Tied	48.8	Tau-a	0.079
Pairs	27778	c	0.619

Wald Confidence Interval for Adjusted Odds Ratios

Effect	Unit	Estimate	95% Confidence Limits	
year_cov	1.0000	2.745	1.641	4.590

Classification Table

Prob Level	Correct Event	Correct Non-Event	Incorrect Event	Incorrect Non-Event	Percentages Correct	Sensi-tivity	Speci-ficity	False POS	False NEG
0.120	86	0	323	0	21.0	100.0	0.0	79.0	.
0.140	61	171	152	25	56.7	70.9	52.9	71.4	12.8
0.160	61	171	152	25	56.7	70.9	52.9	71.4	12.8
0.180	61	171	152	25	56.7	70.9	52.9	71.4	12.8
0.200	61	171	152	25	56.7	70.9	52.9	71.4	12.8
0.220	61	171	152	25	56.7	70.9	52.9	71.4	12.8
0.240	61	171	152	25	56.7	70.9	52.9	71.4	12.8
0.260	61	171	152	25	56.7	70.9	52.9	71.4	12.8
0.280	61	171	152	25	56.7	70.9	52.9	71.4	12.8
0.300	0	323	0	86	79.0	0.0	100.0	.	21.0

Appendix C

C1 Full logistic regression model with all predictors

Table C1 FULL MODEL WITH ALL PREDICTORS

The LOGISTIC Procedure

Model Information

Data Set	COM.ALL
Response Variable	through
Number of Response Levels	2
Number of Observations	398
Model	binary logit
Optimization Technique	Fisher's scoring

Response Profile

Ordered Value	through	Total Frequency
1	THROUGHPUT	85
2	NON-THROUGHPUT	313

Probability modeled is through='THROUGHPUT'.

NOTE: 11 observations were deleted due to missing values for the response or explanatory variables

Class Level Information

Class	Value	Design Variables 1
GENDER	FEMALE	1
	MALE	-1

Model Convergence Status

Convergence criterion (GCONV=1E-8) satisfied.

Model Fit Statistics

Criterion	Intercept Only	Intercept and Covariates
AIC	414.842	390.116
SC	418.828	422.007
-2 Log L	412.842	374.116

The LOGISTIC Procedure

Testing Global Null Hypothesis: BETA=0

Test	Chi-Square	DF	Pr > ChiSq
Likelihood Ratio	38.7262	7	<.0001
Score	38.1826	7	<.0001
Wald	34.0935	7	<.0001

Type III Analysis of Effects

Effect	DF	Wald Chi-Square	Pr > ChiSq
GENDER	1	1.9519	0.1624
african	1	0.2147	0.6431
english	1	2.1264	0.1448
agg_grp	1	4.9836	0.0256
math_grp	1	1.7715	0.1832
immediate	1	2.5019	0.1137
year_cov	1	16.4311	<.0001

Analysis of Maximum Likelihood Estimates

Parameter		DF	Estimate	Standard Error	Wald Chi-Square	Pr > ChiSq
Intercept		1	-2.5021	0.5152	23.5854	<.0001
GENDER	FEMALE	1	-0.1942	0.1390	1.9519	0.1624
african		1	0.2054	0.4432	0.2147	0.6431
english		1	0.5922	0.4061	2.1264	0.1448
agg_grp		1	0.6869	0.3077	4.9836	0.0256
math_grp		1	0.4032	0.3030	1.7715	0.1832
immediate		1	-0.4343	0.2745	2.5019	0.1137
year_cov		1	1.1449	0.2824	16.4311	<.0001

Odds Ratio Estimates

Effect		Point Estimate	95% Wald Confidence Limits	
GENDER	FEMALE vs MALE	0.678	0.393	1.169
african		1.228	0.515	2.927
english		1.808	0.816	4.008
agg_grp		1.987	1.087	3.632
math_grp		1.497	0.827	2.710
immediate		0.648	0.378	1.109
year_cov		3.142	1.806	5.466

The LOGISTIC Procedure

Association of Predicted Probabilities and Observed Responses

Percent Concordant	69.4	Somers' D	0.417
Percent Discordant	27.7	Gamma	0.429
Percent Tied	2.9	Tau-a	0.140
Pairs	26605	c	0.708

Wald Confidence Interval for Adjusted Odds Ratios

Effect		Unit	Estimate	95% Confidence Limits	
GENDER	FEMALE vs MALE	1.0000	0.678	0.393	1.169
african		1.0000	1.228	0.515	2.927
english		1.0000	1.808	0.816	4.008
agg_grp		1.0000	1.987	1.087	3.632
math_grp		1.0000	1.497	0.827	2.710
immediate		1.0000	0.648	0.378	1.109
year_cov		1.0000	3.142	1.806	5.466

Classification Table

	Correct		Incorrect			Percentages				
Prob Level	Event	Non-Event	Event	Non-Event	Correct	Sensitivity	Specificity	False POS	False NEG	
0.040	85	0	313	0	21.4	100.0	0.0	78.6	.	
0.060	85	16	297	0	25.4	100.0	5.1	77.7	0.0	
0.080	78	56	257	7	33.7	91.8	17.9	76.7	11.1	
0.100	77	61	252	8	34.7	90.6	19.5	76.6	11.6	
0.120	73	105	208	12	44.7	85.9	33.5	74.0	10.3	
0.140	67	114	199	18	45.5	78.8	36.4	74.8	13.6	
0.160	66	158	155	19	56.3	77.6	50.5	70.1	10.7	
0.180	63	169	144	22	58.3	74.1	54.0	69.6	11.5	
0.200	53	173	140	32	56.8	62.4	55.3	72.5	15.6	
0.220	51	221	92	34	68.3	60.0	70.6	64.3	13.3	
0.240	49	224	89	36	68.6	57.6	71.6	64.5	13.8	
0.260	44	228	85	41	68.3	51.8	72.8	65.9	15.2	
0.280	36	249	64	49	71.6	42.4	79.6	64.0	16.4	
0.300	36	265	48	49	75.6	42.4	84.7	57.1	15.6	
0.320	30	266	47	55	74.4	35.3	85.0	61.0	17.1	
0.340	27	272	41	58	75.1	31.8	86.9	60.3	17.6	
0.360	19	280	33	66	75.1	22.4	89.5	63.5	19.1	
0.380	18	285	28	67	76.1	21.2	91.1	60.9	19.0	
0.400	17	286	27	68	76.1	20.0	91.4	61.4	19.2	
0.420	12	288	25	73	75.4	14.1	92.0	67.6	20.2	
0.440	12	298	15	73	77.9	14.1	95.2	55.6	19.7	
0.460	11	302	11	74	78.6	12.9	96.5	50.0	19.7	
0.480	11	302	11	74	78.6	12.9	96.5	50.0	19.7	
0.500	11	303	10	74	78.9	12.9	96.8	47.6	19.6	
0.520	5	305	8	80	77.9	5.9	97.4	61.5	20.8	
0.540	5	307	6	80	78.4	5.9	98.1	54.5	20.7	
0.560	5	312	1	80	79.6	5.9	99.7	16.7	20.4	
0.580	5	312	1	80	79.6	5.9	99.7	16.7	20.4	
0.600	5	312	1	80	79.6	5.9	99.7	16.7	20.4	
0.620	0	312	1	85	78.4	0.0	99.7	100.0	21.4	
0.640	0	312	1	85	78.4	0.0	99.7	100.0	21.4	
0.660	0	313	0	85	78.6	0.0	100.0	.	21.4	

Appendix D

D1 Full logistic regression model without year covariate

Table D1 FULL MODEL WITHOUT YEAR COVARIATE

The LOGISTIC Procedure

Model Information

Data Set	COM.ALL
Response Variable	through
Number of Response Levels	2
Number of Observations	398
Model	binary logit
Optimization Technique	Fisher's scoring

Response Profile

Ordered Value	through	Total Frequency
1	THROUGHPUT	85
2	NON-THROUGHPUT	313

Probability modeled is through='THROUGHPUT'.

NOTE: 11 observations were deleted due to missing values for the response or explanatory variables

Class Level Information

Class	Value	Design Variables 1
GENDER	FEMALE	1
	MALE	-1

Model Convergence Status

Convergence criterion (GCONV=1E-8) satisfied.

Model Fit Statistics

Criterion	Intercept Only	Intercept and Covariates
AIC	414.842	406.054
SC	418.828	433.959
-2 Log L	412.842	392.054

The LOGISTIC Procedure

Testing Global Null Hypothesis: BETA=0

Test	Chi-Square	DF	Pr > ChiSq
Likelihood Ratio	20.7879	6	0.0020
Score	21.5206	6	0.0015
Wald	20.2251	6	0.0025

Type III Analysis of Effects

Effect	DF	Wald Chi-Square	Pr > ChiSq
GENDER	1	1.8955	0.1686
african	1	0.7634	0.3823
english	1	2.3136	0.1282
agg_grp	1	7.2464	0.0071
math_grp	1	0.7415	0.3892
immediate	1	1.2009	0.2731

Analysis of Maximum Likelihood Estimates

Parameter		DF	Estimate	Standard Error	Wald Chi-Square	Pr > ChiSq
Intercept		1	-1.9869	0.4812	17.0510	<.0001
GENDER	FEMALE	1	-0.1860	0.1351	1.8955	0.1686
african		1	0.3801	0.4350	0.7634	0.3823
english		1	0.6022	0.3959	2.3136	0.1282
agg_grp		1	0.8117	0.3015	7.2464	0.0071
math_grp		1	0.2545	0.2955	0.7415	0.3892
immediate		1	-0.2899	0.2645	1.2009	0.2731

Odds Ratio Estimates

Effect		Point Estimate	95% Wald Confidence Limits	
GENDER	FEMALE vs MALE	0.689	0.406	1.171
african		1.462	0.623	3.431
english		1.826	0.840	3.968
agg_grp		2.252	1.247	4.066
math_grp		1.290	0.723	2.302
immediate		0.748	0.446	1.257

The LOGISTIC Procedure

Association of Predicted Probabilities and Observed Responses

Percent Concordant	63.0	Somers' D	0.308
Percent Discordant	32.2	Gamma	0.323
Percent Tied	4.8	Tau-a	0.104
Pairs	26605	c	0.654

Wald Confidence Interval for Adjusted Odds Ratios

Effect		Unit	Estimate	95% Confidence Limits	
GENDER	FEMALE vs MALE	1.0000	0.689	0.406	1.171
african		1.0000	1.462	0.623	3.431
english		1.0000	1.826	0.840	3.968
agg_grp		1.0000	2.252	1.247	4.066
math_grp		1.0000	1.290	0.723	2.302
immediate		1.0000	0.748	0.446	1.257

Classification Table

	Correct		Incorrect		Percentages				
Prob Level	Event	Non-Event	Event	Non-Event	Correct	Sensi-tivity	Speci-ficity	False POS	False NEG
0.060	85	0	313	0	21.4	100.0	0.0	78.6	.
0.080	85	4	309	0	22.4	100.0	1.3	78.4	0.0
0.100	85	4	309	0	22.4	100.0	1.3	78.4	0.0
0.120	80	53	260	5	33.4	94.1	16.9	76.5	8.6
0.140	67	62	251	18	32.4	78.8	19.8	78.9	22.5
0.160	63	138	175	22	50.5	74.1	44.1	73.5	13.8
0.180	56	149	164	29	51.5	65.9	47.6	74.5	16.3
0.200	46	210	103	39	64.3	54.1	67.1	69.1	15.7
0.220	42	213	100	43	64.1	49.4	68.1	70.4	16.8
0.240	40	229	84	45	67.6	47.1	73.2	67.7	16.4
0.260	29	233	80	56	65.8	34.1	74.4	73.4	19.4
0.280	27	253	60	58	70.4	31.8	80.8	69.0	18.6
0.300	25	264	49	60	72.6	29.4	84.3	66.2	18.5
0.320	22	268	45	63	72.9	25.9	85.6	67.2	19.0
0.340	16	275	38	69	73.1	18.8	87.9	70.4	20.1
0.360	15	284	29	70	75.1	17.6	90.7	65.9	19.8
0.380	15	289	24	70	76.4	17.6	92.3	61.5	19.5
0.400	8	293	20	77	75.6	9.4	93.6	71.4	20.8
0.420	8	305	8	77	78.6	9.4	97.4	50.0	20.2
0.440	8	307	6	77	79.1	9.4	98.1	42.9	20.1
0.460	0	307	6	85	77.1	0.0	98.1	100.0	21.7
0.480	0	312	1	85	78.4	0.0	99.7	100.0	21.4
0.500	0	312	1	85	78.4	0.0	99.7	100.0	21.4
0.520	0	312	1	85	78.4	0.0	99.7	100.0	21.4
0.540	0	312	1	85	78.4	0.0	99.7	100.0	21.4
0.560	0	312	1	85	78.4	0.0	99.7	100.0	21.4
0.580	0	312	1	85	78.4	0.0	99.7	100.0	21.4
0.600	0	312	1	85	78.4	0.0	99.7	100.0	21.4
0.620	0	313	0	85	78.6	0.0	100.0	.	21.4

Appendix E

E1 Logistic regression - stepwise selection model

Table E1

STEPWISE SELECTION MODEL
The LOGISTIC Procedure

Model Information

Data Set	COM.ALL
Response Variable	through
Number of Response Levels	2
Number of Observations	398
Model	binary logit
Optimization Technique	Fisher's scoring

Response Profile

Ordered Value	through	Total Frequency
1	THROUGHPUT	85
2	NON-THROUGHPUT	313

Probability modeled is through='THROUGHPUT'.

NOTE: 11 observations were deleted due to missing values for the response or explanatory variables

Stepwise Selection Procedure

Class Level Information

Class	Value	Design Variables 1
GENDER	FEMALE	1
	MALE	-1

Step 0. Intercept entered:

Model Convergence Status

Convergence criterion (GCONV=1E-8) satisfied.

The LOGISTIC Procedure

Analysis of Maximum Likelihood Estimates

Parameter	DF	Estimate	Standard Error	Wald Chi-Square	Pr > ChiSq
Intercept	1	-1.3036	0.1223	113.5892	<.0001

Residual Chi-Square Test

Chi-Square	DF	Pr > ChiSq
38.1826	7	<.0001

Analysis of Effects Not in the Model

Effect	DF	Score Chi-Square	Pr > ChiSq
GENDER	1	4.0491	0.0442
african	1	4.2917	0.0383
english	1	5.7137	0.0168
agg_grp	1	14.3227	0.0002
math_grp	1	4.8302	0.0280
immediate	1	1.7478	0.1862
year_cov	1	14.4664	0.0001

Step 1. Effect year_cov entered:

Model Convergence Status

Convergence criterion (GCONV=1E-8) satisfied.

Model Fit Statistics

Criterion	Intercept Only	Intercept and Covariates
AIC	414.842	401.897
SC	418.828	409.870
-2 Log L	412.842	397.897

The LOGISTIC Procedure

Testing Global Null Hypothesis: BETA=0

Test	Chi-Square	DF	Pr > ChiSq
Likelihood Ratio	14.9448	1	0.0001
Score	14.4664	1	0.0001
Wald	13.8429	1	0.0002

Type III Analysis of Effects

Effect	DF	Wald Chi-Square	Pr > ChiSq
year_cov	1	13.8429	0.0002

Analysis of Maximum Likelihood Estimates

Parameter	DF	Estimate	Standard Error	Wald Chi-Square	Pr > ChiSq
Intercept	1	-1.9033	0.2188	75.6663	<.0001
year_cov	1	0.9903	0.2662	13.8429	0.0002

Odds Ratio Estimates

Effect	Point Estimate	95% Wald Confidence Limits	
year_cov	2.692	1.598	4.536

Association of Predicted Probabilities and Observed Responses

Percent Concordant	36.9	Somers' D	0.232
Percent Discordant	13.7	Gamma	0.458
Percent Tied	49.4	Tau-a	0.078
Pairs	26605	c	0.616

Residual Chi-Square Test

Chi-Square	DF	Pr > ChiSq
24.4097	6	0.0004

The LOGISTIC Procedure

Analysis of Effects in Model

Effect	DF	Wald Chi-Square	Pr > ChiSq
year_cov	1	13.8429	0.0002

Analysis of Effects Not in the Model

Effect	DF	Score Chi-Square	Pr > ChiSq
GENDER	1	5.5714	0.0183
african	1	6.1982	0.0128
english	1	6.9592	0.0083
agg_grp	1	13.9054	0.0002
math_grp	1	6.8947	0.0086
immediate	1	3.1519	0.0758

Step 2. Effect agg_grp entered:

Model Convergence Status

Convergence criterion (GCONV=1E-8) satisfied.

Model Fit Statistics

Criterion	Intercept Only	Intercept and Covariates
AIC	414.842	390.499
SC	418.828	402.459
-2 Log L	412.842	384.499

Testing Global Null Hypothesis: BETA=0

Test	Chi-Square	DF	Pr > ChiSq
Likelihood Ratio	28.3427	2	<.0001
Score	27.9470	2	<.0001
Wald	25.8763	2	<.0001

The LOGISTIC Procedure

Type III Analysis of Effects

Effect	DF	Wald Chi-Square	Pr > ChiSq
agg_grp	1	13.4390	0.0002
year_cov	1	13.5511	0.0002

Analysis of Maximum Likelihood Estimates

Parameter	DF	Estimate	Standard Error	Wald Chi-Square	Pr > ChiSq
Intercept	1	-2.2755	0.2530	80.9047	<.0001
agg_grp	1	0.9374	0.2557	13.4390	0.0002
year_cov	1	0.9964	0.2707	13.5511	0.0002

Odds Ratio Estimates

Effect	Point Estimate	95% Wald Confidence Limits	
agg_grp	2.553	1.547	4.215
year_cov	2.709	1.593	4.604

Association of Predicted Probabilities and Observed Responses

Percent Concordant	54.8	Somers' D	0.342
Percent Discordant	20.6	Gamma	0.453
Percent Tied	24.5	Tau-a	0.115
Pairs	26605	c	0.671

Residual Chi-Square Test

Chi-Square	DF	Pr > ChiSq
10.3725	5	0.0653

Analysis of Effects in Model

Effect	DF	Wald Chi-Square	Pr > ChiSq
agg_grp	1	13.4390	0.0002
year_cov	1	13.5511	0.0002

The LOGISTIC Procedure

Analysis of Effects Not in the Model

Effect	DF	Score Chi-Square	Pr > ChiSq
GENDER	1	3.8083	0.0510
african	1	0.6413	0.4233
english	1	2.3402	0.1261
math_grp	1	1.8141	0.1780
immediate	1	3.3946	0.0654

NOTE: No (additional) effects met the 0.05 significance level for entry into the model.

Summary of Stepwise Selection

Step	Effect Entered	Effect Removed	DF	Number In	Score Chi-Square	Wald Chi-Square	Pr > ChiSq	Variable Label
1	year_cov		1	1	14.4664	.	0.0001	
2	agg_grp		1	2	13.9054	.	0.0002	

Wald Confidence Interval for Adjusted Odds Ratios

Effect	Unit	Estimate	95% Confidence Limits	
agg_grp	1.0000	2.553	1.547	4.215
year_cov	1.0000	2.709	1.593	4.604

Classification Table

Prob Level	Correct Event	Correct Non-Event	Incorrect Event	Incorrect Non-Event	Percentages Correct	Sensi-tivity	Speci-ficity	False POS	False NEG
0.080	85	0	313	0	21.4	100.0	0.0	78.6	.
0.100	72	113	200	13	46.5	84.7	36.1	73.5	10.3
0.120	72	113	200	13	46.5	84.7	36.1	73.5	10.3
0.140	72	113	200	13	46.5	84.7	36.1	73.5	10.3
0.160	72	113	200	13	46.5	84.7	36.1	73.5	10.3
0.180	72	113	200	13	46.5	84.7	36.1	73.5	10.3
0.200	61	113	200	24	43.7	71.8	36.1	76.6	17.5
0.220	32	271	42	53	76.1	37.6	86.6	56.8	16.4
0.240	32	271	42	53	76.1	37.6	86.6	56.8	16.4
0.260	32	271	42	53	76.1	37.6	86.6	56.8	16.4
0.280	32	271	42	53	76.1	37.6	86.6	56.8	16.4
0.300	32	271	42	53	76.1	37.6	86.6	56.8	16.4
0.320	32	271	42	53	76.1	37.6	86.6	56.8	16.4
0.340	32	271	42	53	76.1	37.6	86.6	56.8	16.4
0.360	32	271	42	53	76.1	37.6	86.6	56.8	16.4
0.380	32	271	42	53	76.1	37.6	86.6	56.8	16.4
0.400	32	271	42	53	76.1	37.6	86.6	56.8	16.4
0.420	0	271	42	85	68.1	0.0	86.6	100.0	23.9
0.440	0	313	0	85	78.6	0.0	100.0	.	21.4