information generated during the course of a study which has no relationship to the hypotheses of the study? Glaser and Strauss (1971) speculate about what researchers do in this situation. "When discovery forces itself on an analyst, he then writes his induced hypotheses as if they had been thought up before the data were collected, so that they will seem to satisfy the logical requirements of verification (Glaser and Strauss, 1971:185). As Kuhn would say in response to Popper "scientists seek to extend their theories rather than falsify them" (cited in Kneller, 1978:89).

A further problem arises in that the theory which we have accepted as the framework on which to base the present study would not be accepted by Popper as a theory. Especially in its account of ideology, conflict theory has the ability to "save the hypothesis", i.e., to explain apparent refutations of theoretical precepts by means of further reference to the theory. Popper would view conflict theory as "dogma" or "metaphysical theory" (Kneller, 1978:55)

The above problem is unique to the social sciences. In the well-developed natural sciences like physics there is usually one body of theory which is generally accepted and within which all "science" proceeds. Even if there is a transition from one theoretical perspective to another, most

scientists soon end up observing the world guided by a widely accepted theory.

As we have demonstrated in the literature review, the situation with regard to sociology is different. Competing macro-theories co-exist. The researcher's choice of which theory to use when viewing the world is finally, at best, largely a matter of personal choice, and at worst a matter of acquiescing in mob-rule. Indeed, there is much speculation as to how sociologists actually choose their world view (Habermas, 1978). Each body of theory will guide the researcher to look for largely different phenomena in the world. Each theory will also provide the researcher with a different methodological approach to follow when looking for the phenomena (how to find what theory guides you to look for). There is, amongst sociologists, no consensus as to how sociology should be done.

How does the above outline of the problems associated with methodological procedure affect the current study? At an obvious level, it does not seem possible

to escape induction, even though some factors mentioned in the literature review, e.g., financial hardship, lend themselves to prior hypothesis-formation. In the latter instances, however, the conflict theory used to generate the literature review would in principle be able to explain away, with reference to itself, any putative refutation of such hypotheses. Moreover, our research intent is to identify and characterize a novel group of people, and most of this characterization can, we believe, only come from the data itself. In other words we do not want to be blind to what Glaser and Strauss (1971) have termed the unintended consequences of analysis. Indeed, the generation of information is one of the goals of the study. Clearly, hypothetico-deductive orthodoxy is inappropriate in this instance.

Once it is accepted that the approach to the analysis of the data is going to be largely inductive (though from the literature we do have some guidelines) it becomes possible

to apply some common techniques in a different manner, and to use some uncommon techniques of analysis.

Thus, for example, the contingency table and associated probability of association based on the chi square test will be used in what Smith (1975) terms the Dionysian manner: "... the Dionysian knows only the direction in which he wants to go out into the unknown" (Smith, 1975:22). All available variables of possible theoretical importance will be cross-tabulated against both pass/fail and matric rating. The chi square test will be used to differentiate those variables associated with either pass/fail or matric rating from those not associated with either. We will assume that two variables are associated if the probability level arrived at using the chi square test is less than 0.05. We are aware that this means that out of every one hundred "significant" results, five will have occurred by chance.

This danger will be taken into account by the next screening technique we will use - CHAID analysis. (This technique will be explained in more detail below, when it is applied.)

CHAID analysis will be used to further abbreviate the variables associated with pass/fail and matric rating. The

reason for using CHAID rather than some form of factor or cluster analysis is that CHAID is explicitly designed for the analysis of categorical data. Thus, no statistical assumptions about the level of measurement will be violated and we will not have to re-code the data into "dummy" categories for input into, e.g., factor analysis.

Finally CHAID provides the analyst with information on how categories within independent variables behave relative to the dependent variable, i.e., CHAID gives the analyst a suitably stringent method of combining categories within dependent variables so that the newly combined categories do not differ significantly within themselves with respect to the dependent variable. This eliminates the problem of illicit "hunting" for significance by subjectively combining categories of dependent variables using the percentages obtained by cross-tabulating the dependent variable against the independent variable (Rice, 1979).

In the final stage of analysis the remaining re-categorized variables will be entered into a log-linear analysis. (This technique will also be elaborated below, when it is applied.) At this stage no distinction will be made between independent and dependent variables. Log-linear analysis will be used to examine a series of three variable models in

order to determine if there are interactions present between the variables and if all three variables have to be included in the model for it to be adequate.

The hope is that the eventual outcome of the study will be the creation of a simple set of models which will aid people deciding whom to admit to university and whom to exclude. In the words of Sonquist (1977) "hypothetico-deductive rhetoric ... is more fruitfully replaced by a perspective that focuses on multivariate explanatory models which are improved by successive studies" (Sonquist, 1977:325). The overall approach of our analysis to be termed an inductive modelling approach.

The associations we propose to investigate will not, therefore, be stated as hypotheses, but will in Dionysian fashion indicate the general direction the research project will take.

The main thrust of the research will be to examine how the socio-economic and other factors canvassed above in the literature, perhaps mediated by matric performance, affected pass rates at the end of first-year university, using data for B.A. I at Wits in 1982. Some examples of these factors would be whether the respondent is enduring financial

hardship in order to attend university, level of parents education, and race. We will also examine a number of variables which may be either the outcome or the cause of socio-economic position: for example, parents' feelings about university attendance, respondent's feelings about place of residence, feelings about school, number of changes of school, number of standards failed at school, number of people in the respondent's room in matric, whether the respondent's degree course was a first or second choice, and school record from standard eight.

The relationship of a number of other "background" variables to pass/fail will also be examined, such as the respondent's sex, whether his/her parents were divorced or separated or deceased, and the year of matriculation.

3.B. INSTRUMENTS

Ideally social research should proceed according to Denzin's (1970) concept of triangulation. In this view of social research two techniques are used, with the results cross-validating each other. In the case of this study the funds allocated by the University and its need for speedy interim results precluded the use of this ideal technique. We

accordingly chose from the various techniques available the single technique best suited to the study.

In our instance the postal survey was chosen as the most appropriate technique. Caplovitz (1933) points out that as the amount of social research conducted has increased the "population at large" has become unafraid of the schedules of questions which periodically arrive in the post, and tends to accept the guarantee of anonymity.

The result has been an increase in the confidence which social researchers have in the postal survey. As Caplovitz points out: "The assumption of survey researchers is that the great majority of respondents will tell the truth on almost all issues and that if a question is considered too personal, rather than lie, the respondent is likely to refuse to answer" (Caplovitz, 1983:102).

The final version of the questionnaire, which is set out in Appendix A, developed out of a number of drafts and was the end product of the constructive criticism of numerous people involved in the academic support of "disadvantaged" youth at the University of the Witwatersrand.

3.C. METHOD AND ADMINISTRATION

The population for the survey was defined as students registered for first-year Bachelor of Arts at the University of the Witwatersrand in 1982. The questionnaire was administered to all of them with a single follow-up questionnaire some eight weeks later. In order to provide some form of check against respondent bias in answering survey questions about examination results, as much information as possible was also obtained independently from records maintained by the Faculty of Arts. Information obtained from the Faculty included our dependent variable (percentage obtained by the respondent in each first-year final examination) as well as the antecedent variable central to the study (the matric rating of respondents).

The questionnaire was piloted on thirty students involved in the Academic Support Program who all claimed to speak English as a second language. A few minor changes were made to the questionnaire as a result of the pilot.

A master list of all students registered for B.A. I in 1982 was obtained from the Central Data Processing Service of the University. The list took the form of gummed labels which

were used to mail to the respondents envelopes containing a copy of the questionnaire, the covering letter from the Dean and a reply-paid envelope.

As questionnaires were returned the names of respondents were removed from a second copy of the master list. After the eight weeks (at which point no questionnaires had arrived for four days) the remaining labels were used to mail to the defaulters envelopes containing the follow-up letter from the Dean, a further copy of the questionnaire (which had been printed in a different colour to the original) and another reply-paid envelope.

One thousand one hundred and eighty questionnaires were mailed. The initial response was 41%. The response rate increased to 59% after completion of the follow-up, but dropped to 56% once invalid responses had been found and removed (a proportion which McGown (1979) regards as satisfactory for self-administered postal surveys). Invalid responses consisted of questionnaires containing no information except the respondents name, questionnaires returned unopened because of incorrect addresses on the master list, and a few questionnaires with all questions answered in a patently nonsensical manner. Thus, the final working response was 661.

A detailed demographic breakdown of respondents will be provided in Chapter 4. Respondents ranged from seventeen to fifty years old. Ninety eight percent of respondents were under the age of 24. The majority of respondents were female (64%). Reflecting the racial mix of the University, most respondents were White (89%). Indians accounted for 4% of the response, Blacks 2%, coloureds 1%, and Chinese 1%. Four percent of respondents did not indicate to which race group they belonged.

At this stage the Faculty of Arts records were consulted, and the matric ratings of students who had responded to the survey were transferred onto the questionnaires. This information was verified by a second person working independently and all discrepancies were traced and corrected. Students with foreign qualifications and mature-age exemptions were regarded as having "missing values" for the variable.

All information associated with closed-ended questions was then transcribed onto sheets in preparation for card-punching. For the purposes of later verification, the student numbers of respondents was also transferred. The

assignment of numeric responses to open-ended questions took place during the following period of two months.

For each open-ended question (except matric subject registered for) a random sample of one hundred questionnaires was selected. The responses for the question were written down verbatim and a tally was kept of replies. The sheets of transcribed text were then examined and a code frame set up for use on the entire set of responses. Matric subjects were coded according to a schedule provided by the Transvaal Department of Education.

Coding then proceeded. Respondents were allowed two responses for each open-ended question and three responses for the "comments" section. This numeric information was then transcribed onto the card-punch sheets.

Finally the percentage obtained by respondents for each subject registered for at the time of the November examinations was obtained from the Faculty of Arts and transcribed (again with verification).

The data was then captured onto cards and a card image file set up on the University mainframe computer.

Most cross-tabulations and all re-coding of the data was performed using the SAS (1982) package (the program is attached as Appendix B). For the CHAID analysis of Chapter 5 and the log-linear analysis of Chapter 6, files of the relevant variables, suitably categorized, were generated from SAS and used as input into the CHAID package (Kass, 1980) and the BMDP (Dixon, 1983) packages respectively.

4. PRELIMINARY ANALYSIS

4.A. General Frequencies

In order to obtain a "feel" of the data set and to correct punching errors and missing observations, general frequencies were extracted for all variables. Values which were obvious mis-punches or mis-codes were traced to individual records and corrected. Once the data were "clean" a number of preliminary manipulations were performed on some of the variables, to render them amenable to categorical analysis. These manipulations took the form of the collapsing of categories within variables (where appropriate), and the creation of new variables and indices from sets of variables (where necessary).

A summary of the general frequencies and the preliminary manipulations is presented below. The original questions from the schedule are included as an aid.

As mentioned earlier, respondents ranged from seventeen to fifty years old. Ninety eight percent of respondents were under the age of 24. Age was dichotomised, to give as even a split as possible (McGown, 1979): 18 and under (59%) versus 19 and over (41%). More than half the respondents (64%) were female.

The respondents broke down into government race classifications as follows:

Table 4.1 Government Race Classifications

RACE	%
White	89
Indian	4
Coloured	1
Chinese	1
Black	2
No Ans.	4
Total	101
N of cases (661)	

During the course of analysis these categories were combined in a number of ways, to yield composite categories of adequate size for the particular analysis or contrasts of theoretical interest. These are discussed below as they occur.

For each case, the respondent's university subjects and degree course were recorded, and also school performance. The former were found to be relatively unimportant as predictors, the latter very relevant. Use was also made of matric English, Afrikaans and Mathematics symbols (question 1) as predictors.

Rather than constructing specific indices from other matric subjects, (there are some sixty possible matric subjects),

the matric rating of respondents, as provided by the Arts Faculty, was used.

Useful results were also obtained from the analysis of school records prior to matric, where symbols obtained in standards eight, nine and the pre-matric or "mock" matric examination were used.

The matric ratings of respondents occurred in the range 16-64 with a mean of 34 4. During analysis the variable was variously categorised: these arrangements are dealt with as they occur.

Other variables related to school, with their distributions, were:

Table 4.2a Number of Pupils in Matric Class (Qn.7)

NUMBER	%
1 - 23	19
24 - 28	2
29 - 32	23
33 - 45	18
46 +	18
Total	101
N of cases	(657)

Table 4.2b Number of Changes of School (Qn.8)

NUMBER	%
None	70
One	24
Two or More	6
Total	100
N of cases	(648)

Table 4.2c Teacher at All Times in Last Three Years (Qn 9)

RESPONSE	%
Yes	77
No	23
Total	100
N of cases	(654)

For the latter variable, teacher at all times, respondents who were found to have been without a teacher for less than one month were placed in the "yes" category. Matric subjects in which respondents indicated that they were most frequently without a teacher were Physical Science (95%) and Mathematics (73%). The length of time without a teacher varied between two months and a year, with a mean of 4.1 months.

The results dealing with some other questions about school are set out in the following tables.

Table 4.3a Position in Class - Pre-Matric (Qn.6)

POSITION	%
1 - 5	28
6 - 12	38
13 and below	34
Total	100
N of cases (530)	

Table 4.3b Position in Class - Std. Nine (Qn.6)

POSITION	%
1 - 5	29
6 - 12	40
13 and below	31
Total	100
N of cases (522)	

Table 4.3c Position in Class - Std. Eight (Qn.6)

POSITION	%
1 - 5	34
6 - 12	38
13 and below	28
Total	100
N of cases (504)	

Summary results of the remaining questions about school were as follows. Only 7% of respondents had repeated one or more

standards during their school career (question 12). Most felt that they had studied "harder" (26%) or "the same" (39%) as their school fellows (question 15). Respondents were relatively active during their school years in terms of extra-curricular activities (question 15): eight percent indicated that they had not participated in any extra-curricular activity while 70% indicated four or more activities. Sixty-nine percent of respondents stated that they were prefects at school.

Feelings about school attendance (question 14) were grouped into positive (65%), indifferent (17%) and negative (18%). The most common form of positive response were statements like "loved school" or "enjoyed my school years". Respondents who were placed in the "indifferent" category had responses such as "necessary" or "indifferent" to the question. Negative responses included statements like "I hated school" or "felt bad at school". Excluding respondents who were at boarding school, 85% indicated that they shared a room with one person, or had a room to themselves, during their matric year (question 21).

Seventy-seven percent of respondents had been accepted for the degree of their first choice (question 19). Eighty-four percent indicated an intent to study further after their

first degree (question 18). From the question "why did you choose the degree for which you are registered?" (question 18) an index was constructed. The thirteen detailed response themes fitted fairly well into a broader three-way scheme: interest (35%), a career decision (37%), and parental pressure (28%). Indicative examples of interest were "I have always loved people" or "I have always wanted to study psychology"; of career choice, "it is necessary for what I want to do" or "the degree will provide me with a start in business". Parental pressure included statements like "my father said a degree was necessary for me" or "my parents always wanted me to go to university", and "it's important to my parents that I obtain a degree". Twenty-seven percent of respondents felt that someone had particularly encouraged them in their choice of degree (question 5). The most frequently cited person was a parent (52%), followed by friends (11%), and vocational guidance officer (11%). The most common form of encouragement was some form of explanation as to what was involved (30%). This was followed by "setting an example" or "(parent) enjoys occupation so I will" (14%). All other responses occurred in less than 2% of cases.

The feelings of parents, as reported by respondents towards their degree choice (question 6), were generally positive.

Although quite a large number of response categories were generated during coding, these were eventually reduced to positive and negative for the purpose of categorical analysis: fathers 84% and 16% respectively, mothers 89% and 11% respectively. The only notably frequent negative comment generated was an expression of concern about the future career prospects of someone with a B.A. degree (fathers 6%, mothers 4%).

The most common source of finance for studies (question 7) was found to be the respondent's parents (68%). Ten of respondents were totally dependent on on a bursary, scholarship or loan. Twenty-one percent of respondents indicated that their attending university was causing someone financial hardship (question 9).

Eighty-two percent of respondents stated that their home language was English (question 10). No other language was cited in more than 2% of cases. English was also overwhelmingly dominant in conversation with friends and at school, 97% in both cases (questions 11 and 12). Eighty-six percent of respondents were reading non-prescribed books during the university term (question 13). The quantity of books read varied from one to twenty-one with a mean of 3.2.

Fifty-seven percent of respondents read one newspaper daily, 17% none and 26% more than one (question 14).

Seventy-seven percent of respondents were staying at "home", while 14% were staying in university residences (question 17). Thirty-one percent of respondents stated that they would have difficulty in studying at their place of residence (question 19). The most common reasons for students' experiencing difficulty in studying at their place of residence were noise (30%), distractions (17%), and "sharing a room" (13%), (question 19). It is interesting to note here that in the comments section of the questionnaire, 13% of respondents stated that the after-hours library facilities were inadequate or noisy. Travelling time from the place of residence to university varied from five minutes to three hours with a mean of thirty-two minutes (question 20).

Some variables attempting to measure the "quality" of a respondent's home life proved to be important in later analysis. Ten percent of respondents stated that their parents were divorced or separated (question 2). Ninety percent of respondents stated that both their mother and father were still living. Eight percent indicated that their fathers were deceased and 2% that their mothers were

deceased. Only 0.6% of respondents had both parents deceased (question 3).

Parents were the main income earners for 99% of respondents, with fathers at 91%, and mothers at 8% (question 1).

Education of parents fell into the following categories.

Table 4.4 Education Levels of Parents (Qn.4)

EDUCATION LEVEL	FATHER %	MOTHER %
Less than matric	24	30
Matric	30	39
Diploma	13	14
Degree and more	33	17
Total	100	100
N of cases	(597)	(603)

One hundred and sixty different occupations were distinguished among the parents of respondents. These were categorised using an occupational prestige rating scale provided by Castle (1978) and Schmidt (1973). The result is presented below:

Table 4.5 Occupation of Parents (Qn.4)

OCCUPATION CATEGORY	FATHER %	MOTHER %
Category one	22	3
Category two	23	3
Category three	22	16
Category four	8	0
Category five	12	28
Category six	6	1
Category seven	0	1
Category eight	0	1
Category nine	0	0
Category ten	0	0
Category eleven	2	42
Total	100	100
N of cases	(571)	(592)

A summary of the scheme provided by Castle (1978) and Schmidt (1973) is provided below:

CATEGORY — DESCRIPTION

Category 1 - Independent or High Professional. E.g., Doctors, Lawyers, Lecturers, Airline Pilots, Clinical Psychologists.

Category 2 - Salaried Professional and High Managerial and Executive. E.g., Bank Manager, Chief Engineers, Editors, Estate Owners.

Category 3 - Semi-Professional and Lower Executive and Administrative. E.g., Factory Superintendent, Journalists, Nurses Personnel Assistants.

Category 4 - Owners of Substantial Businesses Medium. E.g., Small shops, Garages.

Category 5 - White Collar and Inspectional. E.g., Telephonists, Bank Tellers, Caterers, Representatives, Students.

Category 6 - Skilled Manual, Semi-Skilled Supervisory and Heavy Duty Drivers. E.g., Shunters, Barbers, Soldiers.

Category 7 - Semi-Skilled Manual (Excluding Heavy Duty Drivers) and Unskilled Supervisory. E.g., Unskilled Workers, Apprentices.

Category 8 - Routine Non-Manual. E.g., Bus Conductors, Flower Sellers, "Office Boys", Waiters.

Category 9 - Unskilled Manual. E.g., Factory Hand, Domestics, Loaders.

Category 10 - Menial Laborers. E.g., Mine Workers, Sweepers.

Category 11 - Housewives.

An index of extra-curricular activity for university was calculated (question 15). Students were found to be relatively less active at university than at school (6% participated in no extra-curricular activity, and 40% in four or more activities at university).

From the percentages obtained by respondents at the end of their first year of B.A.,two variables were constructed: the total number of subjects passed, (14% passed 0 subjects, 11% - 1, 15% - 2, 21% - 3, 35% - 4, 4% - 5), and pass or fail (25% fail, 75% pass). The variable pass/fail was constructed according to the Arts Faculty ruling that a student must pass two or more subjects to pass the year.

Category 6 - Skilled Manual, Semi-Skilled Supervisory and Heavy Duty Drivers. E.g., Shunters, Barbers, Soldiers.

Category 7 - Semi-Skilled Manual (Excluding Heavy Duty Drivers) and Unskilled Supervisory. E.g., Unskilled Workers, Apprentices.

Category 8 - Routine Non-Manual. E.g., Bus Conductors, Flower Sellers, "Office Boys", Waiters.

Category 9 - Unskilled Manual. E.g., Factory Hand, Domestics, Loaders.

Category 10 - Menial Laborers. E.g., Mine Workers, Sweepers.

Category 11 - Housewives.

An index of extra-curricular activity for university was calculated (question 15). Students were found to be relatively less active at university than at school (6% participated in no extra-curricular activity, and 40% in four or more activities at university).

From the percentages obtained by respondents at the end of their first year of B.A., two variables were constructed: the total number of subjects passed, (14% passed 0 subjects, 11% - 1, 15% - 2, 21% - 3, 35% - 4, 4% - 5), and pass or fail (25% fail, 75% pass). The variable pass/fail was constructed according to the Arts Faculty ruling that a student must pass two or more subjects to pass the year.

4.8. General Cross-Tabulations and Initial Speculations

In order to evaluate a few associations evident in the literature the constructed variable pass/fail was cross-tabulated with all variables in the data set. In addition matric rating was categorised according to the Arts Faculty rulings in force during 1982 (0-24, 25-34, and greater than 34). This variable was also cross-tabulated with all variables in the data set. A chi square test of significance was calculated for all tables except-two way tables, where SAS uses the Fisher's exact test. The chi square test was chosen as it is a non-parametric technique which is fast (in terms of computer time) and which may be used to establish whether an association exists between two variables or not (Blalock, 1979:299). At this stage of the study this was the only information which was of interest to us. There are measures of the strength of the association between two variables in a contingency table. However, as Upton (1978) remarks "None of these measures should ... be regarded as more than a simple preliminary calculation before a more detailed analysis" (Upton, 1978:38). Since we were certain that almost all the variables selected for further analysis on the basis of their association with our independent variables were going to be re-categorised by CHAID analysis, measures of the strength of association between the

variables were not considered. The information generated during cross-tabulation was also used to uncover any unsuspected associations.

The volume of information produced during this process was huge. What follows is a general discussion of the preliminary cross-tabulations. Probability levels, generated automatically by SAS (1982), are provided when pertinent. Generally, only statistically significant ($p < 0.05$) results will be discussed, except in cases where a non-significant result is meaningful. The tables relating to pass/fail that have been selected for presentation in full here are those dealing with factors which are touched on in the literature review and which are not explicitly tabulated in the later, more detailed analysis.

Age, when initially trichotomised, seemed to be significantly associated with pass/fail (chi square = 48 for 15 d.f., $p < 0.01$). Students older than twenty-four passed in higher proportions than their younger counterparts. However, in the table over 20% of the cells had expected counts of less than five. The same test performed on the variable collapsed into a dichotomy (eighteen and under vs. nineteen and over) did not exhibit statistical significance.

Sex seemed to be significantly associated with pass/fail (chi square = 6 for 1 d.f., p = 0.02), with percentages in the table demonstrating females passing at a higher rate than males. However, examination of the data revealed a significant association between sex and matric rating (chi square = 4 for 1 d.f., p = 0.03). It was evident from an examination of the table that female respondents entering B.A. I possessed higher matric ratings than their male counterparts. The data is presented in Chapter 6 (Table 6.4).

Race was not significantly associated with pass/fail. It was, however significantly associated with matric rating (chi square = 59 for 10 d.f., p < 0.01). This data is also presented in Chapter 6 (Table 6.5).

The actual matric exam, i.e., National Senior Certificate, Joint Matriculation Board or one of the Provincial examinations, written by respondents was not significantly associated with pass/fail. The variable was, however, significantly associated with matric rating (chi square = 24 for 6 d.f., p < 0.01). This data is presented in Chapter 6 (Table 6.6). This follows the pattern demonstrated by the variable race and is to be expected as nearly all the candidates from the Department of Education and Training are

Black (this pattern has been discussed in the literature review).

Most variables referring to symbols obtained during the respondents school career were significantly associated with pass/fail. (In all cases $p < 0.01$.). The cross-tabulations for the various subjects are shown in the following tables. (For each symbol level, the percentage failing is, of course, one hundred minus the percentage passing.)

Table 4.6 Symbols Obtained During School Career by Pass/Fail

MATRIC ENGLISH	% PASS	SAMP. TOT.%
A/B	88	28
C	79	37
D	64	28
E -	56	7
Total		100
N of cases		(626)

chi square = 40 for 3 d.f., $p < 0.01$

MATRIC AFRIKAANS	% PASS	SAMP. TOT.%
A/B	89	16
C	83	28
D	69	36
E -	67	20
Total		100
N of cases		(617)

chi square = 26 for 3 d.f., $p < 0.01$

PRE-MATRIC ENGLISH	% PASS	SAMP. TOT.%
A	91	13
B	76	24
C	81	35
D	57	24
E -	67	4
Total		100
N of cases		(606)

chi square = 41 for 4 d.f., $p < 0.01$

STD. NINE ENGLISH	% PASS	SAMP. TOT.%
A	86	15
B	86	23
C	79	37
D	63	22
E -	58	3
Total		100
N of cases		(651)

chi square = 22 for 4 d.f., $p < 0.01$

MATRIC AGGREGATE SYMBOL	% PASS	SAMP. TOT.%
A	100	4
B	93	12
C	84	36
D	60	41
E -	52	7
Total		100
N of cases		(598)

chi square = 61 for 4 d.f., p < 0.01

PRE-MATRIC AGGREGATE	% PASS	SAMP. TOT.%
A	100	4
B	96	16
C	77	42
D	64	31
E -	53	7
Total		100
N of cases		(583)

chi square = 52 for 4 d.f., p < 0.01

STD. NINE AGGREGATE	% PASS	SAMP. TOT.%
A	97	5
B	90	20
C	75	44
D	64	29
E -	47	3
Total		101
N of cases		(602)

chi square = 39 for 4 d.f., p < 0.01

STD. EIGHT AGGREGATE	% PASS	SAMP. TOT.%
A	90	10
B	82	25
C	71	45
D	65	19
E -	42	2
Total		100
N of cases		(598)

chi square = 23 for 4 d.f., p < 0.01

From the above tables it can be seen there is a clear association between symbols obtained at school, from standard eight to matric, and passing at university. If one looks at the percentages in the tables it can be seen that this association becomes weaker as one moves back to the less recent symbols. This is what one would expect. The

trend is also apparent in the following tables concerning maths.

Table 4.7 Matric and Pre-Matric Maths Symbols by Pass/Fail

MATRIC MATHS	% PASS	SAMP. TOT.%
A	88	6
B	92	6
C	88	17
D	77	26
E -	64	42
Total		99
N of cases		(577)

chi square = 35 for 4 d.f., $p < 0.01$

PRE-MATRIC MATHS	% PASS	SAMP. TOT.%
A	85	7
B	92	11
C	79	21
D	66	34
E -	66	34
Total		100
N of cases		(551)

chi square = 23 for 4 d.f., $p < 0.01$

Mathematics symbols obtained in standard nine and in standard eight were also significantly associated with pass/fail (for mathematics symbol in standard nine - chi square = 26 for 4 d.f., $p < 0.01$; for mathematics symbol in standard eight - chi square = 18 for 4 d.f., $p < 0.01$).

Respondents showed quite strong continuity of performance during their school careers. Generally, those who performed well in standard eight continued to do so to through matric and university. The "ad hoc" measures relating to "position in class during school career" were all significantly associated with pass/fail (for position in class in matric

chi square = 33 for 2 d.f., $p < 0.01$; for position in class in standard nine - chi square = 21 for 2 d.f., $p < 0.01$; for position in class in standard eight - chi square = 17 for 2 d.f., $p < 0.01$). Once again respondents showed a reasonable degree of consistency through their school careers.

The number of students in the respondents' matric class was not significantly associated with pass/fail, or with matric rating. It is possible that a very high number of students in the matric class does affect matric performance, but that on the basis of their consequently poor matric marks a high proportion of these people are excluded from university (and by default from our study). The variable "number of people in respondent's room during matric was dichotomised into the categories "did not share a room" versus "did share a room". The variable was not significantly associated with pass/fail. However, the variable was significantly associated with matric rating (chi square = 7 for 2 d.f., $p = 0.02$).

The number of changes of school that a respondent underwent was significantly associated with pass/fail, as shown in the table below:

Table 4.8 Number of Changes of School by Pass/Fail

NO. SCHOOL CHANGES	% PASS	SAMP. TOT.%
None	78	70
One	69	24
Two +	74	6
Total		100
N of cases		(633)

chi square = 6.019
for 2 d.f, p = 0.049

Looking at the percentages within the table it can be seen that the association did not present a linear trend as might have been expected. Rather, those students who underwent no change of school during their school career performed better than those with one school change, who in turn performed worse than those who changed schools two or more times. A likely speculation is that students who are in the third category are determined "survivors". The variable was not significantly associated with matric rating.

The presence of a teacher at all times was not significantly associated with pass/fail. The variable was associated with matric rating (chi square = 6.4 for 2 d.f., p = 0.04). The length of time scholars were without teachers was not significantly associated with either pass/fail or matric

rating. Repeating a standard at school was associated with pass/fail.

Table 4.9 Repetition of a Standard at School by Pass/Fail

REPEAT STD.	% PASS	SAMP. TOT.%
None	77	93
One	52	5
Two +	75	2
Total		100
N of cases		(657)

chi square = 14 for 2 d.f., p < 0 01

Looking at the percentages within the table it can be seen that those who had repeated a standard at school tended to fail in higher proportions than those who had not.

The students' subjective rating of how hard they studied in comparison to their school fellows was significantly associated with pass fail.

Table 4.10 Index of Relative Work in Matric by Pass/Fail

RELATIVE WORK MATRIC	% PASS	SAMP. TOT.%
Harder	81	29
Same	76	40
Less	69	31
Total		100
N of case		(645)

chi square = 7.19
for 2 d.f., p = 0.03

Once again, if the percentages within the table are looked at, a trend is apparent. Those who rated themselves as hard workers performed better than those who rated themselves as working "the same", and both categories performed better than students who rated themselves as working less than their colleagues. This variable was also associated with the matric rating of respondents (chi square = 26 for 4 d.f., p < 0.01).

Whether students were studying their second or first degree choice was not significantly associated with their pass rate. This seems rather paradoxical since there is an association between degree choice and matric rating (chi square = 6.3 for 2 d.f., p = 0.04). From the percentages in the table it could be seen that respondents doing a "second choice" degree have lower matric ratings than their "first

Table 4.10 Index of Relative Work in Matric by Pass/Fail

RELATIVE WORK MATRIC	% PASS	SAMP. TOT.%
Harder	81	29
Same	76	40
Less	69	31
Total		100
N of cases		(645)

chi square = 7.19
for 2 d.f., p = 0.03

Once again, if the percentages within the table are looked at, a trend is apparent. Those who rated themselves as hard workers performed better than those who rated themselves as working "the same", and both categories performed better than students who rated themselves as working less than their colleagues. This variable was also associated with the matric rating of respondents (chi square = 26 for 4 d.f., $p < 0.01$).

Whether students were studying their second or first degree choice was not significantly associated with their pass rate. This seems rather paradoxical since there is an association between degree choice and matric rating (chi square = 6.3 for 2 d.f., p = 0.04). From the percentages in the table it could be seen that respondents doing a "second choice" degree have lower matric ratings than their "first

choice" follows. The intent of respondents to study further after their first degree was not significantly associated with their pass rate. A possible explanation is that it was still too early in the student's university career for the variable to be a strong motivating factor.

Students' reasons for registering for their degree was significantly associated with the variable pass/fail.

Table 4.11 Reasons for Degree Choice by Pass/Fail

REASONS FOR DEGREE CHOICE	% PASS	SAMP. TOT.%
Interest	81	34
Career	71	37
Parents	78	29
Total		100
N of cases		(588)

chi square = 6.8 for 2 d.f., p = 0.03

Looking at the percentages it can be seen that students who had registered as a result of some form of "interest" performed the best followed by those with some form of parental pressure. Respondents whose reasons for choice of degree were classified as "career decision" performed the worst. A clarification is necessary. Parental pressure is not a negative category. From the open-ended responses it was clear that what occurred was often more than strong

encouragement; but there were, as far as could be interpreted, no cases of active coercion In contrast to the above finding, students' feelings about attending school was not significantly associated with their pass/fail rate. The variable was however significantly associated with matric rating (chi square = 15 for 4 d.f., $p < 0.01$).

Feelings of parents towards degree choice was not ...ificantly associated with pass/fail. The source of finance for the students study was not quite significantly associated with pass/fail (chi square = 10 for 5 d.f., $p = 0.06$). Looking at percentages in the cross-tabulation it was evident that students who were dependent on a bursary, loan or scholarship and students supporting themselves tended to pass in greater numbers than other students. The explanation is (for once!) clear. Students obtaining bursaries, loans and scholarships tend to have better matric marks than remaining students. Self-supporting students tend to be much older than their counterparts.

The variable establishing whether university attendance was causing anyone "financial hardship" was significantly associated with pass/fail.

Table 4.12 University Attendance Causing Financial Hardship by Pass/Fail

FINANCIAL HARD.	% PASS	SAMP. TOT.%
Yes	62	21
No	79	79
total		100
N of cases		(638)

chi square = 16 for 1 d.f., $p < 0.01$

The variable was also significantly associated with matric rating (chi square = 11 for 2 d.f., $p < 0.01$). The finding supports the view stated in the literature review, that financial hardship is one manifestation of a culturally deprived environment (Parker and Kleiner, 1971), i.e., students might be doing less well at school not because of the direct affect of financial hardship but because of other factors which are manifestations of financial hardship, for example, comparatively low self-esteem.

The question establishing whether students would have difficulty studying at their place of residence was significantly associated with pass/fail.

Table 4.13 Difficulty in Studying at Place of Residence by Pass/Fail

DIFFICULT STUDYING RESID.	% PASS	SAMP. TOT.%
Yes	69	31
No	78	69
Total		100
N of cases		(650)

chi square = 5.6
for 1 d.f., p = 0.02

Looking at the percentages within the table it is evident that those who replied "yes" had a higher failure rate than those who answered "no". This finding is important in the light of comments on library facilities which were described earlier. Neither the reason for study difficulty nor the place of residence significantly associated with pass/fail. Travelling time, using uncollapsed categories, was however significantly associated with pass/fail (chi square = 38 for 25 d.f., p = 0.05). Due to the sparse distribution of travelling time, over twenty percent of cells had expected values of less than five. The p value of the variable (collapsed into ten categories) was 0.06 (chi square = 16 for 9 d.f.) - not quite statistically significant.

The variables "parents divorced", "parents deceased", "home language" and "school language" were not significantly

associated with pass/fail. All were, however, significantly associated with matric rating (for parents divorced - chi square = 9 for 2 d.f., p = 0.01; for parents deceased - chi square = 14 for 6 d.f., p = 0.03; for home language - chi square = 50 for 30 d.f., p = 0.01, for school language - chi square = 27 for 14 d.f., p = 0.03). Looking at the tables it was clear that respondents who indicated that their parents were divorced or deceased were clustered in the "low" end of the matric rating scale. Respondents whose home or school language was not English also tended to fall into the "low" end of the matric rating scale.

The occupation of parents was not significantly associated with pass/fail or matric rating. The level of parents education was, however, significantly associated with pass/fail, as shown below:

Table 4.14 Parents' Level of Education by Pass/Fail

	MOTHER		FATHER	
LEVEL OF EDUCATION	% PASS	SAMP. TOT.%	% PASS	SAMP. TOT.%
Lt matric	72	30	68	23
Matric	74	40	72	31
Diploma	74	14	70	12
Degree	89	16	85	34
Total		100		100
N of cases		(610)		(610)
chi square = 11 for 3 d.f., p = 0.01			chi square = 15 for 3 d.f., p < 0.01	

The education level of the mother was significantly associated with matric rating (chi square = 13 for 6 d.f., p = 0.02). Interestingly, the education level of the father was not significantly associated with matric rating.

From the above table it can be seen that having a parent with a degree placed respondents at a distinct advantage.

The general pattern that emerges from the cross-tabulations was that some variables were significantly associated with both pass/fail and matric rating, while other variables were significantly associated with only the one or the other. A table summarizing this pattern, for the most pertinent variables, is presented below:

Table 4.15 Variables Significantly Associated With Matric Rating or Pass/Fail or Both

VARIABLE	PASS/FAIL ONLY	MATRIC RATING ONLY	BOTH
Sex			*
Race		*	
Matric exam written		*	
Number of people in room in matric year		*	
School changes	*		
Teacher at all times		*	
Repetition of a standard	*		
Working harder or same			*
Feelings about attending school		*	
Financial hardship			*
Divorce of parents		*	
Home language		*	
School language		*	
Education of father	*		
Education of mother			*
Variables related to school			*

This pattern and its implications will be examined in detail later.

There are a number of dangers inherent in attempting to associate matric rating at the same time as pass/fail with a given set of attitudinal variables. At the time the survey was administered students had not yet written their final exams at university. However, their matric marks were (of course) known to them and it is possible that some form of rationalization with regard to the latter had taken place. These thoughts and explanations would obviously have been reflected in the attitudinal questions.

In addition the matric ratings of respondents are biased upwards as many potential respondents at the lower end of the matric rating scale were excluded from the University.

In the preliminary associations considered above, it was seen in several instances that a two-way association might well be qualified in some way by a third. This is a well-known problem in survey analysis. Yet to control each relationship by all other antecedents would have been a vast and aimless task. So, to try and clarify the composite picture all variables significantly associated with

pass/fail and matric rating were selected for input into CHAID (Kass, 1980) analysis.

5. CHAID ANALYSIS

The computer program, CHAID (Kass, 1980) was used for further analysis of the abbreviated variable set considered in the previous chapter. All variables included for CHAID analysis were either theoretically important, in the sense of arising from the literature, or significantly associated with pass/fail or matric rating.

According to Hawkins and Kass (1982) the original function of automatic interaction detection (AID) routines (on which CHAID, the categorical version, is based) was to "discover whether the linear hypothesis model predicting Y from X contains only main effects or whether interactions also occur" (Hawkins and Kass, 1982:269). In the present context CHAID was not actually used for this purpose. As Hawkins and Kass (1982) point out, AID techniques have other uses which they consider might overshadow the initial function of AID. DuToit (1983) also explains some important "side effects" of CHAID.

The most important subsidiary use of CHAID in the present analysis was the merging of categories within predictor (independent) variables. These "new" categories differ significantly from each other in terms of the predicted

(dependent) variable. However, within themselves the merged categories of the independent (predictor) variables do not differ significantly with respect to the dependent (predicted) variable.

The power of this process of collapsing categories is enhanced by means of various "classes" or types of predictor which are present within CHAID. The programme distinguishes three main classes of predictor variables:

1. Monotonic Predictors. The categories within these predictors are considered ordered, e.g., matric marks. If this form of predictor is specified *CHAID will not combine* categories across other categories. For example, in the variable matric marks, category A will not be combined with category E - only adjacent categories will be combined.
2. Floating Predictors. If the predictor is specified as "floating" roughly the same procedure applies. The predictor is considered ordered with the exception of one category *which may be combined with any other category*. The implication is, of course, that this option copes with missing data, grouping missing cases where they would "most likely be" with that category from which they are least significantly different. The program may of course leave

missing cases as a separate category if they are unique in their relation to the predicted variable.

3. Free Predictors. As the name implies no restrictions are placed on the combination of categories within these predictors, which would thus usually be nominal variables such as race, choice of degree, etc.

The second use made of CHAID is found in DuToit (1983). CHAID was used to screen a set of variables already considered "relevant" in order to reduce this set to a group of variables which might be termed "important". This set of variables could then be inserted into, say, a log-linear model which would be the final form of analysis.

Full descriptions of CHAID analysis including examples can be found in Hawkins and Kass (1982). DuToit includes in his 1983 paper a clear description of how to complete a CHAID run.

Fundamentally, CHAID splits predictor variables associated with a predicted variable into categories. The predictor variable most significantly associated with the predicted variable is first selected for splitting. "The X statistic and Bonferroni levels of significance are used to make the split" (DuToit, 1983:5). Bonferroni levels of significance

take into account the problem of completing numerous tests of significance simultaneously. This solves the problem mentioned on p. 80 that, if one does a hundred tests while working with a significance level of $p < 0.05$, an average of five will come up "significant" just by chance. Associations are now required to have a much higher p value than 0.05 before they are regarded as significant. Since the p value will alter depending on how many variables are included in the analysis we will term significant results as "significant at Bonferroni $p < 0.05$". The actual p value will generally be much lower.

The initial split gives rise to the first level of prediction. For example, say we are attempting to analyse education in the categories (high)/(low) against a set of variables, e.g., income (high)/(medium)/(low)/(undisclosed), sex (male)/(female) and area (rural)/urban). The program will scan the variable set and select the variable which is most Bonferroni-significant with respect to education, having previously tested all combinations of the categories of the independent (or predictor) variables in order to obtain the optimal "split" for each one vis-a-vis education. Let us suppose that the predictor variable income is related to the predicted variable education in two "splits", (high, undisclosed),(medium, low). The program is informing the

analyst that, for differentiating the two levels of education, income is the best predictor, and that the income categories (high)/(undisclosed) are not significantly different from each other with respect to education. (Similarly the income categories (medium)/(low) are not significantly different from each other with respect to education.) The four-category variable income may therefore be collapsed into a two-category variable - income (high)/(low). There will be a significant difference between the two categories with regard to level of education. Each of the splits within income will in turn become a predicted variable, i.e., income (high) will be the next predicted variable, followed by income (low), etc. CHAID will search the remaining predictor variables for the variable most significantly related to any or each of the two splits. The output from CHAID analysis may be represented as a "tree" diagram, e.g., from our hypothetical variable set:

Author Classen Gregory Mark
Name of thesis Identifying Disadvantage: Some Antecedents Of Academic Performance Among First Year Arts Students At The University Of The Witwatersrand In 1982. 1987

PUBLISHER:
University of the Witwatersrand, Johannesburg
©2013

LEGAL NOTICES:

Copyright Notice: All materials on the University of the Witwatersrand, Johannesburg Library website are protected by South African copyright law and may not be distributed, transmitted, displayed, or otherwise published in any format, without the prior written permission of the copyright owner.

Disclaimer and Terms of Use: Provided that you maintain all copyright and other notices contained therein, you may download material (one machine readable copy and one print copy per page) for your personal and/or educational non-commercial use only.

The University of the Witwatersrand, Johannesburg, is not responsible for any errors or omissions and excludes any and all liability for any errors in or omissions from the information on the Library website.