University of Cape Town

School of Economics

Youth Employment in the Cape Town Area

Insights from the Cape Area Panel Study

Mamello Lekena
LKNZAN001

Mini-Dissertation in partial completion of Masters in Economics

Supervisors: Prof. Murray Leibbrandt
Prof. David Lam

Abstract

In South Africa today, one of the most urgent economic, social and political challenges is the extraordinarily high and increasing level of unemployment. While the economy has been growing steadily, unemployment levels have continued to rise. A large part of the problem is the fact that the number of new entrants outstrips the creation of new jobs (Ashton, 2005; Chichelo et al., 2003). Since most of the new entrants are people making the transition from school and other education institutions into the labour market, it is not surprising that the problem of unemployment is more severe among the youth. Young people as a group are disproportionately affected by unemployment – whereas the unemployment level for adults is in the region of 26%, the rate for youth is 50%, while 58% of the unemployed are young people (Mlatsheni & Rosphabe, 2002). This paper looks at the factors that affect employment among the youth in the Cape Town area. Using data from the 2002 and 2004 waves of the Cape Area Panel Study, cross-sectional and panel probit regressions have been employed to investigate the role of individual, household and schooling characteristics in the probability of employment. The results suggest that although household income and the presence of employed people in the household were important in a cross-sectional setting, previous labour market experiences were more important in a dynamic setting. While quantity of schooling is found to be very important, the impact of quality of schooling could not be determined. The usual race and gender patterns in employment were found.

Plagiarism Declaration

1. Plagiarism is to use another's work and try and pass it off as my own. I know that this is unacceptable.

   This paper entitled "**Youth Employment in the Cape Town Area - Insights from the Cape Area Panel Study**" is my own work.

2. I have used the convention for citation and referencing adopted by *The South African Journal of Economics*. Each significant contribution and quotation in this paper from the work(s) of other authors has been attributed and has been cited and referenced.

3. I have not allowed, and will not allow anyone to copy my work for plagiarism purposes.

Mamello Lekena (LKNZAN001) Signed by candidate
July 2006

Table of Contents

1. Introduction

In South Africa, a middle-income developing country, one of the most urgent economic, social and political challenges is the extraordinarily high and increasing level of unemployment. Unemployment levels are so high that the accuracy of the estimates has been questioned by international authors and organisations (Wittenberg, 2002). According to Statistics South Africa, the current unemployment rate is 26.5%, using the narrow definition. Under this definition, the unemployed are defined as those who have actively searched for work in the two weeks prior to the survey (Statistics South Africa, 2005). Even using this narrow definition and excluding all those who, for whatever reason have not actively searched for work, the numbers are alarming. This mass unemployment has also meant that there is a large gap between the narrow (official) definition and the broad or expanded definition of unemployment (Wittenberg, 1999). Kingdon & Knight (2000) find that the difference between the narrow and expanded unemployment rates in South Africa from 1993 to 1997 was never below 11% and was, at one point even as high as 18.2%. This means that up to 40% of the labour force is unemployed under the broad definition. It also means that there are millions of South Africans who have not searched for employment but who insist that they want to work.

A common notion that has often come up in discussions on unemployment is the phenomenon of 'jobless growth' of the South African economy. By this, it is meant that the South African economy, which has shown modest growth over the past 10 years, has nonetheless been shedding jobs over the period. While it is true that unemployment levels have been rising, it is incorrect to conclude that the economy has failed to create jobs. In the period 1995 – 2002, 1.6 million new jobs have been created in the economy. The reason for the increased level of unemployment is that over the same period, 5 million people entered the labour market. This means that 3.4 million of them have become or remained unemployed (Bhorat, 2004).

A large part of the problem is thus the fact that the number of new entrants outstrips the creation of new jobs (Ashton, 2005; Chichelo et al., 2003). Since most of the new entrants are people making the transition from school and other education institutions into the labour market, it is not surprising that the problem of unemployment is more severe among the youth. Young people as a group are disproportionately affected by unemployment – whereas the unemployment level for adults is in the region of 26%, the rate for youth is 50%, while 58% of the unemployed are young people (Mlatsheni & Rosphabe, 2002). There are a host of other factors that contribute to this figure. Along with the large size of the youth labour force, lack of experience and less (accurate) information about job prospects also contribute to youth unemployment. The combination of low education and lack of experience that is characteristic of a large proportion of the new entrants means that young people experience more difficulty in finding employment when they enter the job market (Mlatsheni & Rosphabe, 2002).

Also contributing to increasing levels of unemployment is the fact that the growth of the world economy in general has been characterised by a shift in focus from primary sectors to the tertiary or services sector (Ashton, 2005). The South African economy has also followed this trend, and although agriculture and mining continue to play an important role, there has been a shift towards the services sector which has been characterised by significant increases in capital-labour ratios (Bhorat, 2004). This has, predictably, led to an increase in the demand for highly skilled workers, and a gradual slump in the demand for their lower-skills counterparts. Output growth is, and is expected to continue to be skills-biased. For example, in manufacturing, the proportion of skilled workers increased from 6% to 10% while the proportion of unskilled declined from 19% to 15% in the period from 1995 to 2002. In the same period, the proportion of unskilled workers in mining fell from 19% to 7% (Bhorat, 2004).

Given South Africa's history, it is fairly intuitive that high skills jobs were previously reserved for Whites while the rest of the population were relegated to low-skills manual labour. In an effort to correct these imbalances and to catch up and keep up with the world economy, the present government has made it a national priority to move the labour force from a low-skills equilibrium to a high-skills equilibrium (Ashton, 2005). A National Skills Development Strategy (NSDS) was formulated as the key policy tool in this effort. The NSDS however, focuses mainly on improving skills levels of people who are already employed in low-skills jobs, and therefore doesn't directly impact on new entrants to the job market, and particularly youth. It has become increasingly important for young people to obtain skills before entering the labour market in order to be able to secure and maintain employment.

With young people entering the job market in their numbers and the ongoing redistribution of jobs from low to high skills, it is clear that those entering the job market with higher skills levels will be at an advantage. The importance of post high school education cannot be overestimated. Wittenberg (2002) has shown that the impact of education on the probability of employment is most significant at the post-matric level. In a comparison of employment levels by education level, he found that the relation is rather flat (and below 100%) at all education levels and then increases sharply to around 100% for people with post-matric diplomas and degrees. It is thus clear that people with higher education qualifications have less difficulty in finding employment than their less educated counterparts. At the same time, one cannot ignore the fact that most of the young people entering the labour market do not have post matric qualifications. In fact a large number of people leave school before obtaining a matric qualification and, for a variety of reasons, enter the job market. Although it is understandable that low-skills jobs tend to have negative political and historical connotations, the fact is that jobs need to be found to accommodate these people if the youth unemployment challenge is to be dealt with in an effective fashion. It is therefore critically important to understand the factors that affect employment of young people without higher education qualifications.

This paper attempts to provide some answers to the question posed above. Given the macroeconomic backdrop outlined here, we look at the micro characteristics and constraints that affect access to jobs for those without higher education qualifications. The question that we attempt to answer therefore is – what are the characteristics that determine which young people, particularly among those without post-matric qualifications, manage to find employment? Having already established that in Sub-Saharan Africa, the youth unemployment rate is even higher than the alarmingly high unemployment rate for the whole population, it is critical that we come to terms with the causes and determinants of employability, particularly for youth. Young people also form a much larger component of the labour market in Africa than in other areas, making the problem of youth unemployment a key labour market and development issue (Leibbrandt and Mlatsheni, 2004).

The paper proceeds with an outline of the South African and international literature on unemployment in general, and youth unemployment in particular in section 2. Section 3 gives an overview of the Cape Area Panel Study. Section 4 discusses some descriptive statistics for the 2002 wave of CAPS, and section 5 continues with a description of the 2002/2004 panel and a discussion of some potential econometric problems that are associated with panel data. Section 6 presents the results of the cross section analysis of the determinants of the probability of employment while section 7 covers the labour market transitions and the results of the analysis of employment in a dynamic setting. Section 8 concludes.

2. An overview of the literature.

The microeconomic causes of unemployment in general and youth unemployment in particular, have been the subject of much analysis and discussion in economic literature. In South Africa, it is unfortunate but still true that employment varies drastically by race, gender and location. For example, the youth unemployment rate in 1996 was 40.9%, but that of women was 49.6% compared to 33% for men. Rural youth face higher unemployment rates and the unemployment rate of African youth is the highest at 50.2% (Sekwati & Hirschowitz, 2001). South Africa's history has made race and gender very salient factors in any labour market discussion. Age is also important when analysing youth unemployment because many African youth make the transition from school into the labour market very late as many of them are still in the schooling system into their twenties. When they do eventually enter the job market, they find it more difficult than White youth to find work. By age 25, almost all white males are employed, while only 50% of African males have employment (Wittenberg, 2002).

The characteristics of the household in which a young person resides also have an influence on that person's probability of finding employment. Kingdon & Knight (2000) found that the unemployed live in much poorer households and are more likely to live in rural areas. 70% of the unemployed in South Africa live in households that have less than R800 expenditure per month. This means that most of the unemployed cannot afford to spend money on searching activities (Bhorat, 2004). The labour market status of the adults in a household has a bearing on the labour market outcome of the young people in the household. There is a direct effect due to the fact that the employment and earnings of the adults affect the amount of resources available to fund education, migration and job search of the young people (Leibbrandt & Mlatsheni, 2004). Poor families tend to have larger numbers of unemployed people who aren't even searching for work because the cost of searching is prohibitive.

In addition to this, there is the effect of networks and their economic impact on the way people interact. The broad theory is that the people that one knows and engages with have an impact on one's decisions and behaviour and thus an impact on one' economic and social outcomes. A person's networks include family, friends, neighbours and anyone that the person has any contact with. The effect of the network on an individual's decisions is two-fold, an information channel and a social norm channel (Bertrand et al, 2000). In the case of job searching, the most important of these is the information channel – people will wait for employed people within their network to inform them of available job opportunities. One of the strongest predictors of a person being employed is the presence of another employed person within their household. If a parent is unemployed, the child is likely to be unemployed too, (Wittenberg, 1999). It follows that the presence of any employed person in the household will be beneficial to the labour market outcomes of the young people in that household. This information channel is particularly important when the household cannot afford to fund conventional job search

activities and the young people in that family have to rely on their networks for relevant job information.

In labour economics literature, the education level and income of the household head are often used as proxies for the labour market status of all the adults in the households, and as a measure of the socio-economic status of the household (Leibbrandt & Mlatsheni, 2004). A young person living in a household where the household head has a low level of education, is more likely to unemployed that a similar youth with a highly educated household head. A similar case can be made for the effect of the income level of the head of the household. In this paper, we will use these and other characteristics to analyse the effect of the household on young people's labour market outcomes.

Arguably, the most important factors that are emphasised in labour market literature are investments in human capital – schooling and skills training (Becker, 1964; Mincer, 1974). As argued above, higher education levels almost automatically translate to better prospects in the labour market. For youth, this effect is even more pronounced, as young people with lower education levels are likely to have more difficulty in securing employment that their older counterparts (Mlatsheni & Rosphabe, 2002). While it is encouraging to find that the historical gaps in years of education achieved across racial lines has decreased (Anderson & Lam, 2003), it is generally accepted that the *quality* of the education received is just as important a factor as the quantity of education in determining employment and earnings (Van der Berg, 2001).

Although there is some debate around the most accurate way to measure quality of education, one can generally take an inputs approach and look at expenditure on schooling per child, teacher-pupil ratios, infrastructure and resources available at school and so on. Alternatively, one could look on the output side, which would mean taking into consideration the students' performance in some standardised test. Both methods are imperfect, but the do shed some light on the issue hand. Tyler et al. (1999) found that in the US, cognitive skills are an important factor in the employment of high school drop-outs. Holzer (1996) considers a range of skills, including mathematical and linguistic skills, and concludes that even on the demand side, employers are reluctant to employ people who do not possess these skills. Of course it isn't clear whether there are returns to cognitive skills directly, or if the market is rewarding some other characteristic that is correlated to numeric and literacy skills – but these are still good determinants of employment (Holzer, 1996). Chamberlain and Van der Berg (2002) come to similar conclusions by including literacy and numeracy scores as measures of quality of education. They found that accounting for quality of education diminishes the wage gap between races – indicating that discrimination by race happens prior to entry into the labour market. In other words, most the what one may see as labour market discrimination in employment and earnings by race, is in fact a function of the vast differences in the quality of the schooling that was offered to different racial groups. This means that because poor African households were more likely to be subjected to poor primary schooling, even though they may achieve the same number of years of

schooling than their White counterparts, their labour market outcomes are nonetheless still compromised (Van der Berg, 2001).

The challenge of youth unemployment is not new, and has been studied from many different angles in the literature. Over and above the extensive international literature on the subject, the problem of youth unemployment in South Africa has received considerable attention. The closest analysis to this one is probably the work done by Mlatsheni & Rosphabé (2002). In that paper, they also look at the micro determinants of youth unemployment, and contrast youth unemployment to unemployment in the rest of the labour force. This paper differs from others on the subject in three important ways. Firstly, the focus in this paper is exclusively on youth. We use data from the Cape Area Panel Study, a study of 4700 or so young people in the Cape Town area. The survey is specifically a survey of youth in the area, and data is collected from respondents who were 14 – 22 years of age in 2002. This means that we have detailed information, from the youth perspective. In previous papers that analysed youth unemployment using data from the 1993 PLSDS (for example, Wittenberg & Pearce, 1996), the October Household Surveys (OHS) or the Labour Force Surveys (LFS), it was difficult to obtain information on parental background and other details pertaining to intergenerational transfer of inequalities (Mlatsheni & Rosphabé, 2002). In the CAPS dataset, this information is captured explicitly.

The second important distinction is the introduction of detailed schooling data, *at the individual level*. It is often difficult to obtain detailed schooling data, especially in the developing country context (Van der Berg & Burger, 2002). Using data from the Western Cape Department of Education's School Register of Needs (SRN), we have access to data on the schooling experience of each respondent, based on the school that they actually attended, rather than, for example, aggregate data from schools within their neighbourhood. This data is used to look at the impact of schooling on probabilities of employment. The schooling data is further complemented by Literacy and Numeracy Evaluation (LNE) scores based on a literacy and numeracy test which was administered as part of the CAPS survey. Since each respondent in the survey has a LNE score, we have a direct measure of their literacy and numeric skills. This is, for example, in contrast to Chamberlain & Van der Berg (2002) who used test scores from the 1993 PLSDS to impute test scores for their OHS (1998) respondents. While these LNE scores may not be a perfect measure of literacy and numeracy skills, and the disadvantages of using test scores have been explained at length in the literature (eg Heckman et al (2003), Chamberlain & Van der Berg (2002)), they serve as a useful and well measured indicator of quality of school output.

The third and most important distinction comes from the use of panel data, which allows us to track the progress of individuals over time. A similar methodology was used by Dinkelman (2002), but while she analyses searching and unemployment using KIDS data, we use CAPS data to analyse youth unemployment. The major advantage of CAPS in this regard is that the data is very recent and the gap between the waves is two years as opposed the five-year gap in KIDS. Using panel

data allows us to see the effect of micro characteristics on unemployment in a dynamic setting. We will use the panel to analyse the impact of 2002 characteristics on 2004 outcomes, thus giving a more nuanced picture of the functioning of the youth labour market.

Although our dataset is restricted to the Cape Town area, which some may argue is not typical of the rest of the country, the richness of the data clearly compensates for the limited geographic span of the study. The Cape Area Panel Study is described in greater detail in the next section, along with a description of the School Register of Needs.

3. The Cape Area Panel Study

In this paper, we use data from the Cape Area Panel Study (CAPS), a project that is a collaboration project between the University of Cape Town (UCT) and the University of Michigan. CAPS is a detailed longitudinal survey of young people between the ages of 14 and 22 in the Cape Town area, and their families. The first wave was conducted in the latter half of 2002, with wave 2 conducted in 2005. In each of the intervening years, a mini-wave was conducted, which collected information from a subset of the respondents. The second of these mini-waves was conducted in 2004 and captured more detailed education and labour market information from *two thirds* of the 2002 respondents. We use this mini-wave as the second wave for the purposes of this paper. In 2002, in addition to a household questionnaire, a Young Adult questionnaire was administered, which collected detailed information about schooling, employment, fertility and other demographic data. This was accompanied by a life calendar which was used to collect retrospective data on schooling, living arrangements and sexual partners. This data has been linked to data on school quality from the Western Cape Department of Education's School Register of Needs (SRN). In this way, we have information on school quality at the individual level, making this survey a unique one for South Africa and allowing us to complete the analysis by adding the effect of schooling characteristics on the relevant outcomes. In 2002, respondents were requested, in addition to completing the survey, to take a basic Literacy and Numeracy Evaluation (LNE), the results of which are also used in this study.

The CAPS survey uses a two-stage probability sample of households. Approximately half of the population of Cape Town is made up of coloured people, with white and black/African people making up 22% and 27% of the population respectively. Because we desired roughly equal numbers of young adults from each racial group, we over-sampled African and white households relative to coloured households. The first-stage sample of Census Enumeration Areas (EAs) was drawn using the 1996 Census as a sampling frame. Households within EAs were sampled directly from 1998 aerial ortho-photos which clearly showed the residences within each selected EA. EAs likely to have experienced substantial population change since the last enumeration (e.g., EAs identified by the Census as containing a high proportion of informal settlements) were visited on the ground by teams trained to obtain updated listings. Using this updated sampling frame, a sample of households was drawn for the interview teams to visit.

Once a household was contacted for interview, a decision on whether to include it in the survey was made based on the household's demographics. All households containing at least one resident between the ages of 14 and 22 were selected for inclusion in the sample. Upon recruitment into the survey, the household demographic questionnaire was administered to the person most knowledgeable about the household. Full-length young adult interviews were given separately to up to three young adults in the household. Up to five visits were made to each household to make these interviews.

Based on the data from the 1996, a sample was drawn that was anticipated to provide completed interviews with 2,652 African, 2,314 coloured and 2,293 white households. These households were projected to give 2,119 African, 1,856 coloured and 810 white young adults. The final sample contains 2,072 African, 1,989 coloured and 727 white households (and 28 Indian/Other), and 2,032 African, 1,838 coloured and 557 white young adults (and 26 Indian/Other). The number of African and coloured households interviewed was slightly less than anticipated, but nearly as many African and coloured young adults as expected were obtained because household size in African and coloured areas appears to have increased since 1996. Non-participation rates for whites were higher than we had anticipated. The baseline wave of CAPS provides data on 4,816 households and 4,453 young adults (Anderson & Lam, 2003; CAPS Technical Documentation).

4. Descriptive statistics for the 2002 wave of CAPS

The CAPS data contains information about respondents who were 14 – 22 years of age in 2002. For the purposes of the analysis in this paper, we look at those youth who have completed 12 or less years of education, that is high school or less, and who are no longer enrolled in school. This group includes people who have completed Matric and have not continued with tertiary education, as well as people who have dropped out before completing high school. While there are people in the survey who have incomplete tertiary education, there are not enough of them to be able to draw statistical conclusions from. Another small group of respondents have complete tertiary education, but they are not a homogenous group – their qualifications vary from 6 month certificates to four year degrees. While it would be beneficial to draw comparisons between this group and the more vulnerable group that are the focus of this paper, the small number of such observations as well as their heterogeneity, make it difficult to analyse these comparisons in a statistically meaningful way.

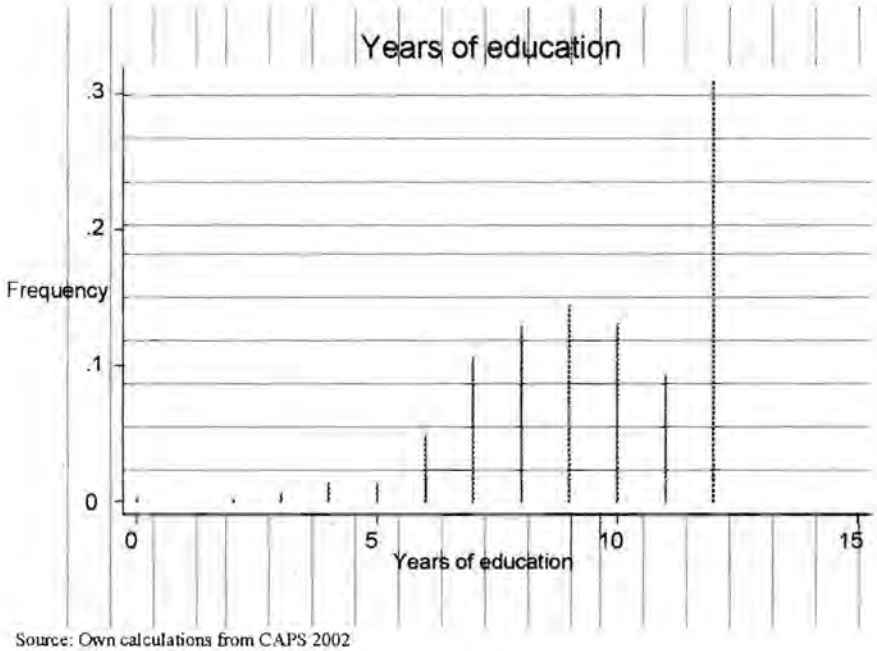
Having thus restricted the sample, we have 1627 young people in our 2002 sub-sample. Of these, there are 707 African[1], 852 Coloured, 2 Indian and 66 White respondents. Moreover, there are 48.6% or 730 male respondents and 51.4% or 897 female respondents. However, the sampling technique used in CAPS and the higher than expected non-participation rates, as mentioned in the previous section, necessitate the use of weights in order for our sample to give a better representation of the underlying population that we are studying. In this study, we use analytical weights for this purpose. Therefore, having weighted the sample, we find that the racial composition represented by the sample is 443 African, 1088 Coloured, 2 Indian and 92 White. The gender composition changes slightly to 791 male and 836 female respondents. It is clear that one cannot justifiably infer anything of substance from our sample about Indians in Cape Town, since there are only two Indian respondents in the sample. For the remainder of the analysis therefore, we will not draw any conclusions about the characteristics and outcomes of Indians.

Figure 1 (overleaf) shows the frequency of years of education for the sample under investigation in 2002. It therefore represents youth (14-22 years) who are no longer enrolled in formal education and who have not completed more than 12 years of education. Given South Africa's high primary school enrolment rates, it is not surprising to see that there are very few people who have incomplete primary education. Although there are about 5% who have only completed 6 years, the first significant spike is at 7 years of education, which represents complete primary education. Approximately 14% of the people in the sample have completed Grade 9, which is the highest level of compulsory education in South Africa, and would be a

[1] Throughout this paper, we will use the term "African" to refer to Black Africans, as opposed to Coloured or White respondents.

kind of natural exit point. The most prominent spike is at 12 years of education. 30.98% of the people in the sample have completed high school. While the large spike is encouraging, there are still large numbers of young people who are exiting the school system prematurely. Given the fact that the distribution of years of education shown in figure 1 above shows no evidence of being quadratic in nature, it is unnecessary to include a squared education variable in the regression analysis in the later sections of the paper.

Figure 1: Years of education of youth who have completed at most 12 years of schooling



Source: Own calculations from CAPS 2002
Note: The data have been weighted

Table 1: Education and employment characteristics of young adults (ages 14-22), by population group

	School Enrolment		Employment	Years of education	
Race	Still in school	Not in school	Employed	Mean	Std Dev.
African	888	443	77	9.205	2.362
%	66.74	33.26	17.52		
Coloured	1356	1087	517	9.540	2.200
%	55.51	44.49	47.55		
White	698	93	68	11.451	1.175
%	88.27	11.73	73.72		
Total	**2976**	**1625**	**664**	**9.559**	**2.253**
%	**64.69**	**35.31**	**40.94**		

Source: Own calculations form CAPS 2002
Notes:

1. Employed refers to people who have worked in the 12 months prior to the survey.
2. The data have been weighted.

The first part of table 1 below shows the proportion of youth who have 12 or less years of education, who are still enrolled in school. A total of 64.69% are enrolled. That means that 35.31% of the young adults are not in school, and these are the young adults that we will analyse in the rest of the paper. From Figure 1, it is clear therefore that only 30.98% of these 31.35% have completed high school. This means that of these 35.31% who are not enrolled, 69.02% have dropped out prior to finishing their high school education. The racial decomposition is reflective of South Africa's socio-economic dynamics. 88.27% of White youths are enrolled, while only 66.74 of African youths are in school. The case of Coloured youths is even more dire, with just over half (55.51%) enrolled. One of the factors that could be driving these disparities is that most White youths continue to further education after completing Matric, while large proportions of African and Coloured youths exit the education system at this point. The underlying factors that determine the progress of young people from high school to tertiary education range from financial to educational constraints, but are beyond the scope of this study.

Given the worldwide shift from primary to tertiary industries (Ashton, 2002) and the increasing demand for highly skilled labour at the expense of low-skilled labour (Bhorat, 2004), it is almost inevitable that these young people with low levels of education will be hard put to find employment in the current labour market. The second two columns of table 1 refer to the youths who have exited the schooling system. Almost 60% of them (59.06%) are unemployed, an unemployment rate that is about 20% higher than the national broadly defined unemployment rate. The picture is more nuanced when one looks at the racial breakdown of the employment rates among this group. White youth have a far higher than average employment rate at 73.72%. Even though Coloured youth have an employment rate that is higher than the average for the whole group, at 47.55% employment, it is still much lower than the national rate. African youths have the lowest incidence of employment, with a mere 17.52% of them employed. Alarming though this may be, it is hardly surprising and it is a clear reflection of the lingering effects of Apartheid on the labour market.

The final column in table 1 shows the average years of education by race. The average schooling follows the same trend as the unemployment rate. White youth have an average of 11.45 years of education. This means that a considerable number of them have in fact completed high school. Coloured youths have, on average 9.54 years of education and African youths have the lowest average level of education at 9.2 years. This may in fact be one of the factors driving the patterns of unemployment as shown in the second column of table 1.

Table 2: Education and employment characteristics of young adults, by gender

	School Enrolment		Employment	Years of education	
Gender	Still in school	Not in school	Employed	Mean	Std Dev.
Male	1412	789	389	9.282	2.376
%	64.13	35.87	49.33		
Female	1565	835	275	9.820	2.098
%	65.20	34.80	32.99		
Total	**2976**	**1625**	**664**	**9.559**	**2.253**
%	**64.69**	**35.31**	**40.94**		

Source: CAPS 2002

Notes:

1. Employed refers to people who have worked in the 12 months prior to the survey.
2. The data have been weighted.

Table 2 contains the same analysis as table 1, but decomposed by gender. As far as enrolment goes, there isn't much difference in the enrolment rates of males and females. Males have an enrolment rate of 64.13% while the females have a slightly higher rate at 65.2%. Employment rates however, are very different across gender. Half of the males are unemployed, compared to 67% of the females. This is, counter-intuitively, not reflective of the average years of education, where the females have more years of education (9.82 years) than the males who have 9.28 years.

Household characteristics

The literature review suggests that the type of household that one resides in is key in determining probabilities of employment. Access to household income is crucial, both for acquiring human capital and for funding job search activities. The household is also crucial in providing access to labour market information networks through household members who are employed (Bhorat, 2004). Furthermore, the presence of young children or elderly people in the household may impose care-giving responsibilities which limit entry into the job market. At the same time, the social security grants provided for children (Child Support Grant) and for the elderly (State Old Age Pension) add to overall household income, and may therefore be beneficial in assisting job search. Given the supply-side focus of this paper, it is therefore important to understand the nature of the households that the youths under investigation live in, so that we can establish the effect of these household characteristics on the employment prospects of the youth.

The educational background of the head of the household or parent has an impact on the labour market experience of a young person, in direct and indirect ways. The "silver-spoon, plastic-spoon" hypothesis suggests that children in families with higher levels of education, are themselves likely to have higher levels of education (Burns, 2001). Furthermore, the higher the education level of the head of the household, the more likely they are to be employed, and therefore able to impart valuable and more accurate labour market information. Along with employment, higher education may also be linked to higher household income (Western et al., 1998), which would alleviate some of the financial constraints that

may hinder the acquisition of higher education and the funding of job-search activities.

As a first measure of household characteristics, we look at the characteristics of the head of the household. Table 3 gives some of these characteristics, broken down by race.

Table 3: Characteristics of the household head, by population group

	Highest grade completed		Proportion with tertiary qualification		Proportion employed	
Race	**Mean**	**Std Dev.**	**Mean**	**Std Dev.**	**Mean**	**Std Dev.**
African	7.684	3.209	0.017	0.130	0.629	0.483
Coloured	7.853	2.838	0.081	0.273	0.667	0.471
White	11.258	1.071	0.391	0.492	0.774	0.422
Total	**8.016**	**2.987**	**0.082**	**0.275**	**0.663**	**0.473**

Source: CAPS 2002

Notes:

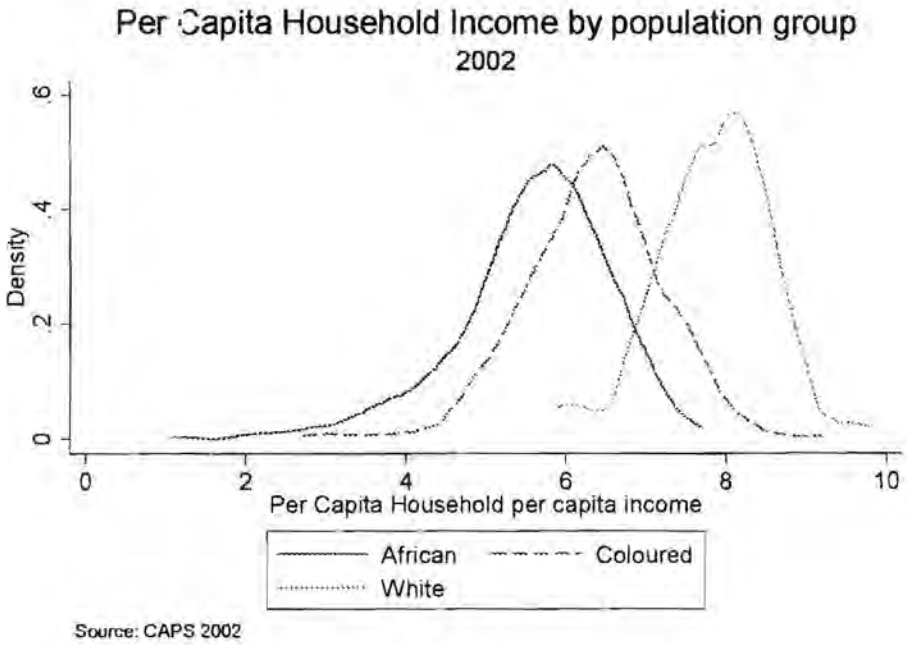
1. Employment refers to the proportion of people who have worked in the 12 months prior to the survey.
2. Tertiary qualification is any post-Matric qualification. This ranges from certificates and diplomas to degrees and post-graduate degrees.
3. The data have been weighted.

The average years of education of household heads across the sample is 8 years, which means just one year of high school education. Again, a racial breakdown of this figure is telling. African and Coloured household heads have, on average, 7.68 and 7.85 years of education, which is barely higher than a primary school qualification. White household heads have a mean of 11.26 years of education. There is much wider variance in the education levels of African and coloured household heads than is the case for their White counterparts. This could, hopefully, be an indication that although the average education levels in non-white households are low, there are some households that are moving up the socio-economic ladder with regards to education. The proportion of household heads with a tertiary qualification follows a similar pattern, only in this case the disparity is more extreme. Close to 40% of White household heads have a tertiary qualification, whereas only 8% of Coloured and a mere 1.7% of African household heads have a post-Matric qualification. This is of course related to the fact that African and Coloured household heads have much lower average years of education. In fact most of them have not completed high school, which would automatically mean that they cannot have a tertiary qualification. That being said, one should bear in mind the finding by Wittenberg (2002) that employment levels are almost consistently 100% for people with post-Matric diplomas and degrees. This means that the African and Coloured household heads are less likely to be employed that their white counterparts. This will clearly have an impact on household income levels and other socio-economic indicators that will affect the performance of young people in the labour market.

The final column of table 3 shows the proportion of household heads that have been employed in the 12 months prior to the survey. At first glance, the racial

differences are not as severe as one would expect, given the findings in the previous two columns of the same table. 62% of African household heads and 66.7% of coloureds are employed. This is not very far behind the 77.4% employment level of White household heads. What the table does not mention however, is the quality of employment that is referred to. It is very feasible to assume that White household heads would have more stable and higher paying jobs, especially given their superior educational backgrounds. The data does indeed show that most White household heads have full time jobs while almost 20% of African household heads are only employed on a part-time basis. Figure 2 shows the estimated densities of log per capita household incomes of the three population groups[2]. This is the total income from all sources, including non-labour income. Per capita income has been used to standardise across households of varying size.

Figure 2: Per capita household income



Source: CAPS 2002

Note: Data have been weighted

African households have the lowest incomes, with white households at the other extreme. Their density functions only intersect at the extremes, meaning that there is very little overlap in the income levels. Coloured households' incomes are in between the two other population groups. The mean per capita income of African households is R384, compared to R857 for Coloured households and R3129 for White households. These patterns fit in with the findings regarding the education and employment characteristics of household heads, even though this analysis takes into account all the employed people in a particular household and not just the head of the household head.

[2] The kernel density functions have been estimated using the standard assumptions.

Education characteristics

Having highlighted the importance of human capital in the labour economics arena, and noted the intuitive links between education and the labour market, we look at some of the key education indicators in the CAPS dataset. As mentioned earlier, one of the highlights of this dataset is that because we are able to link the CAPS data to the School Register of Needs (SRN), we have access to schooling information at the individual level. We also use information about schooling from the CAPS dataset. Although this is really the impressions of the respondent of his or her school, it does provide valuable insight. Looking at these school characteristics gives an input approach to measuring school quality. The other side of the coin would be to look at an output approach. Towards this end, we use results from the Literacy and Numeracy Evaluation that was administered as part of the CAPS survey.

Table 4: Schooling characteristics, by population group

	Nr of school problems		Pupil-Teacher Ratio		LNE Literacy scores		LNE Numeracy scores	
Race	Mean	Std Dev.	Mean	Std Dev.	Mean	Std Dev.	Mean	Std Dev.
African	1.646	1.601	34.170	5.076	58.479	18.092	30.147	20.307
Coloured	2.852	2.603	30.819	3.625	70.262	13.654	43.475	24.230
White	1.014	1.329	24.334	4.749	80.159	10.166	74.286	20.057
Total	2.417	2.396	31.001	4.483	67.616	16.005	41.652	25.045

Source: Own calculations from CAPS 2002 and School Register on Needs (SRN) 2000
Note: The data have been weighted.

The first column of table 4 gives the average number of problems that the respondents experience at school. In the CAPS survey, the young adults were asked to select from a list of 13 problems, which they had ever experienced at their school. Since the learners in our sample are no longer in school, this refers to the most recent school that they attended. The problems listed range from a shortage of textbooks to drug dealing and sexual harassment[3]. The creation of a single variable to capture all of these problems does not by any means suggest that the problems are trivial or that they have an equal impact on a learner's experience. Another important caveat is that these figures are obtained from the CAPS data which means that they are based on what the respondents reported, and have not been verified by the school authorities. They are, nonetheless an indication of the differences in schooling experiences by race. White respondents reported the least number of problems, with an average of 1 problem reported. Coloured students experienced almost three times as many problems, with African learners falling in between the other two groups.

[3] The list of problems that the respondents selected from was: Not enough textbooks, Dirty classrooms, Crowded classrooms, Teacher often absent from class, Teachers drunk, Teachers being threatened by students, Noisy classrooms, Drug dealing, School environment unsafe, Sexual harassment by students, Sexual harassment by teachers/staff, Bullying by students and Violent teachers

The pupil-teacher ratio is often used as a measure of the quality of education. It gives some indication of the resources dedicated to each child. There may, of course be cases where a high pupil-teacher ratio is mitigated by better quality teaching. White youth have an average of 24.33 students per teacher. This stands in stark contrast to their African peers' learning experience of 34.17 pupils per teacher. Coloured youth have an average pupil-teacher ratio of 30.82. While this measure makes it is clear that there is still a considerable racial divide in access to quality education, it is not ideal for the purposes of this study. The pupil-teacher ratio is, to some extent standardised across schools, and there is therefore little variation across individual schools. As a result, we use a measure of the proportion of governing body appointed teachers to government appointed teachers as a similar measure of quality of education. The Department of Education permits schools that have the means to, to appoint additional teachers over and above the standard allocation of teachers. These additional teachers are remunerated by the governing body of the school which means that schools will only appoint as many teachers as their resources allow, with poorer schools having no governing body teachers and private schools having only governing body appointed teachers and no state funded teachers.

Moving to an output based measure of quality of education, we look at test scores. In this particular literacy and numeracy evaluation (LNE), respondents were given the option of completing the evaluation in English or Afrikaans. 99% of the Xhosa speaking respondents chose to complete the LNE in English. This means that for these people, the LNE was an evaluation of basic English as well as literacy and numeracy. It is clear therefore that comparing LNE performance across population groups may be less than optimal. We proceed with the necessary caution.

The test scores recorded in table 4 have been standardised as percentages. White and Coloured youth outperform African youth in both the literacy and the numeracy tests. This is very likely to be influenced by the languages in which the LNE was administered, as mentioned above. That being said, the difference in the scores of White and Coloured respondents, both of which took the LNE in their first language, is striking, and confirms the patterns of quality of education shown by the input based measures. What is also remarkable is the difference, within each population group, between the literacy and the numeracy scores. In all three cases, the numeracy scores are much lower than the literacy scores – with Coloured youth showing the biggest gap of close to 27%. This may speak to the quality of mathematical education across the board, but one can clearly not draw explicit conclusions form this basic analysis in that regard.

5. The panel data – descriptive statistics and data issues

As mentioned earlier, we have divided the full CAPS dataset into various sub-samples. The CAPS sample is designed to be a representative sample of the population of Cape Town youth in 2002. The main sub-sample, those who were no longer enrolled in 2002 and had completed 12 or less years of schooling, was chosen because we made the key assumption that they are a group of particular interest in the study of labour markets. The fact that they're not enrolled in any form of education, have not got any tertiary qualifications and are now competing in the labour market, makes them a particularly vulnerable group, and the focus of this paper. It is therefore important to see how this group compares to a representative sample of Cape Town youth, namely the full CAPS sample. Table 5 compares some of the key characteristics of the various sub-samples that have been drawn from the full CAPS sample.

Table 5: Descriptive Statistics for the various sub-samples in 2002 and 2004

	2002 Full Sample	2002 valid sample	2002/2004 Panel in 2002	2002 Sub-sample not interviewed in 2004	Valid in 2004	2002/2004 Panel in 2004
CAPS Wave	*2002*	*2002*	*2002*	*2002*	*2004*	*2004*
N	4740	1627	1123	504	1200	1123
Mean Age	17.92	19.61	19.55	19.76	20.55	21.41
% Female	52.41	51.40	49.55	55.39	50.01	49.19
Mean Education	9.49	9.56	9.50	9.69	9.84	9.49
% with Matric	17.78	30.98	28.84	35.58	34.13	28.76
% African	28.25	27.24	27.28	27.16	23.39	23.34
% Coloured	52.35	66.89	69.25	61.81	71.96	74.11
% White	18.65	5.71	3.37	10.76	4.44	2.39
% Employed	23.42	40.94	40.19	42.55	47.27	45.49
Mean per capita Household Income	1401.00	843.31	756.65	1032.21	-	-
Mean education of Household Head	8.92	8.02	7.75	8.58	-	-
Mean education of Mother	8.29	6.74	6.57	7.10	-	-
Mean education of Father	6.71	4.92	4.78	5.23	-	-

Source: CAPS 2002 and CAPS 2004
Notes:
1. All data except for sample sizes are calculated using sample weights.
2. Household information is not available in the 2004 wave of CAPS

The first column of table 5 refers to the full CAPS sample. The second refers to the sample of youth who, in 2002, were not enrolled and had at most a grade 12 qualification, which will be referred to as the 2002 valid sample. This group represents 34% of the full sample. Some of the differences between the two samples (age, mean education, employment) are due to the fact that the latter consists mostly of people who are still enrolled in high school. There are fewer Whites in the 2002 valid sample (5.71% compared to 18.65%). This is probably because White youth are less likely to drop out before Matric and more likely to continue to tertiary education. The household variables show a big gap between the full sample and the 2002 valid sample. The latter group are from poorer households and the adults in their households have lower levels of education on average. There is therefore evidence to show that the group that we have selected for this study are a more vulnerable group than the general population of youth in Cape Town.

One of the key problems associated with panel data is attrition bias. From one wave of a survey to the next, it is often not possible to trace all the respondents from the previous wave. It is also possible that respondents from one wave may simply refuse to be re-surveyed. If this attrition is systematic, then parameter estimates may be inconsistent (Dinkelman, 2002). The same problem can also occur because of selection bias. This is potentially true of the CAPS survey, where the 2004 sample consists of only two thirds of the full sample that was surveyed in 2002. Although this was largely by design, it is important to establish the extent to which the group that was re-surveyed in 2004 differs from the 2002 group.

The third column of table 5 refers to the respondents who were in the 2002 valid sample, who are also in the 2004 wave of the survey. The fourth refers to those who were in the 2002 valid sample, but not in the 2004 wave. Ideally the second and third columns should be as similar as possible. In other words, one would like the sub-sample of respondents who were re-interviewed in 2004 to be as similar as possible to the sample from the first wave. The sub-sample in column 3 is the one that is used for the panel regressions. Most of the key characteristics are very similar across the two waves. There are fewer Whites and thus a larger proportion of the sample is Coloured. From Column 4, it is clear that there were more Whites who were not included in the 2004 wave. The per capita income of the 2004 group is lower, and this may be due to the fact that there are fewer Whites in the sample. The education variables for the adults in the household are very similar. Although there are some small differences between the data from the two waves, it seems safe to assume that the effects of selection bias, if any, will be minimal.

The second common problem associated with panel data is measurement error. While measurement error is a problem for most, is no all surveys, panel datasets are more hard-hit. One way of checking for measurement error in this case is to compare the sample in column 3 to the sample in the last column of table 5. The last column refers to the same sample as column 3, but shows the data recorded from the 2004 wave. Ideally, the two columns should be identical, save for age, which should have a two year difference and employment, which could change in any direction. The differences in the gender and the mean education variables point to two things. The first is fairly accurate measurement, and secondly to the fact that not many of the respondents from the 2002 sample retuned to formal education between 2002 and 2004.

The fifth column of table 5 refers to the sample that is used for the 2004 cross-sectional regressions. This sub-sample is selected in the same way as the sample used fro the 2002 cross-sectional regressions (valid 2002 sample). That is, it consists of those respondents who were not enrolled in 2004 and had at most a matric qualification in 2004, referred to as the 2004 valid sample. Although this group will certainly include respondents who were in 2002 valid sample, it will include others who were enrolled in 2002 and were no longer enrolled in 2004, and exclude those from the 2002 valid sample who may have returned to school by 2004.

As expected, the average age is higher, although not as high as the two year gap in the waves would indicate. The mean education is slightly higher and the proportion with matric is also higher at 35.58%. This is consistent with the fact that the 2004 sample is older and will therefore include more people with higher levels of education. The percentage employed is also higher in 2004, which may be related to the higher education levels among the 2004 sub-sample.

Having had a look at some of the key variables and characteristics of the sample, we proceed in the next section to evaluate the impact of these and other variables on the probability of employment for youth in the Western Cape. We begin by looking at cross section analyses of both the 2002 and the 2004 waves before proceeding to the dynamic analysis across both waves.

6. The results of the cross-section analysis

The results of the probit estimations for the 2002 and 2004 cross-sections are presented in table 6. The dependent variable (employed) is a binary variable that is 1 if the respondent has worked for pay in the 12 months prior to the survey, and zero otherwise. Coefficients are represented as the marginal effects, evaluated at the mean for continuous variables, and evaluated as the discrete change from 0 to 1 for dummy variables. The co-efficients therefore give the percentage change in probability of being employed for a given change in the variable of interest, when all others are evaluated at their means. Z-statistics are reported in brackets.

Personal Characteristics

Column 1 shows the impact of individual level variables only, for 2002, while column 2 shows the same analysis for 2004. As expected, personal characteristics matter in determining the probability of employment. All the coefficients have the expected signs and most are significant at the 5% level. Being female has a negative effect, while being Coloured or White (as opposed to Black, which is the reference group) has a positive effect. The impact of education is not as large as one would expect, with an additional year of education only increasing the probability of employment by 2.4% in 2002 and 2% in 2004. Having a matric qualification matters more in 2004, with matriculants being 21.4% more likely to be employed. In 2002, the corresponding figure is only 8.9%. The other individual characteristics have a similar impact on employment across both waves of the survey.

Household characteristics

Column 3 shows the result of the regression including household characteristics. Household income, which is measured net of any income generated by any employed young adults in the household, could have two opposing effects on the probability of employment. On one hand, higher income means that there are more resources available in the household that can be used to facilitate search activities and thus lead to a higher probability of employment. A potential problem here is that the household income is related to the number of employed people in the household, which is also included as an explanatory variable. In the analysis however, we find that the coefficient of the income variable doesn't change when we include the proportion of employed adults (number of employed adults as a proportion of the total adults in the household) in the regression. Higher income could also result in an income effect where young adults from wealthier households have les pressure to work, and therefore choose to consume more leisure. However, given the extremely high unemployment levels in South Africa, and given that unemployment is even more severe among the youth, this type of argument is not as readily applicable in this context as it might otherwise be.

The marginal effect of household income is positive and significant. A 1% increase in household income increases the probability of employment by 21.5%. This confirms the first hypothesis that higher household income results in a higher probability of employment.

Table 6: Determinants of employment among youth – 2002 and 2004 cross-section probit regression results (Marginal effects)

	(1) 2002 Individual variables	(2) 2004 Individual variables	(3) Household Variables	(4) Schooling (LNEs)	(5) Schooling (SRN)
	2002	2004	2002	2002	2002
Individual Variables					
Female	-0.191	-0.156	-0.176	-0.167	-0.197
	(6.48)**	(4.46)**	(5.31)**	(4.93)**	(5.04)**
Coloured	0.322	0.455	0.256	0.264	0.327
	(11.41)**	(13.43)**	(7.65)**	(7.43)**	(6.91)**
White	0.48	0.416	0.288	0.277	0.454
	(6.28)**	(4.59)**	(2.61)**	(2.48)*	(3.52)**
Education 2002	0.024		0.024	0.024	0.035
	(2.30)*		(2.85)**	(2.41)*	(3.26)**
Matric 2002	0.089				
	(1.92)+				
Age 2002	0.053		0.049	0.049	0.054
	(6.18)**		(5.19)**	(5.07)**	(4.99)**
Education 2004		0.02			
		(1.58)			
Matric 2004		0.214			
		(3.91)**			
Age 2004		0.044			
		(5.36)**			
Household Variables					
Log of HH Income			0.215	0.213	0.212
			(7.95)**	(7.79)**	(6.54)**
Female Head			-0.046	-0.044	-0.047
			(1.38)	(1.32)	(1.22)
Education of Head			-0.012	-0.009	-0.016
			(1.83)+	(1.45)	(2.15)*
Head with Tertiary qualification			0.089	0.081	0.125
			(1.14)	(1.04)	(1.41)
Proportion Employed			-0.094	-0.093	-0.082
			(4.78)**	(4.65)**	(3.62)**
Nr of pre-school children			0.067	0.071	0.069
			(3.29)**	(3.44)**	(2.85)**
Nr of elderly people (age>60)			-0.08	-0.076	-0.065
			(2.33)*	(2.21)*	(1.67)+
Schooling Variables					
Pupil-Teacher Ratio					0.012
					(2.32)*
Private Teachers					0.111
					(0.29)
LNE Literacy Score				-0.008	
				(1.52)	
LNE Numeracy score				0.006	
				(1.28)	
Observations	1621	1182	1434	1411	1043

Source CAPS 2002 and CAPS 2004

Notes: 1. Robust z-statistic in brackets

2. + significant at 10%; * significant at 5%; ** significant at 1%

3. The log of household income and proportion employed variables are net of the young adult's income and employment

While this may be due to a number of factors, it is very likely to be related to resources available for searching, especially since income here is not restricted to employment income, but includes income form social grants, which form a large proportion of the income in low income households.

The presence of employed adults in the household is expected to have a positive impact on the probability of being employed. Employed people within the household constitute a link to the labour market by being able to provide the young adults in the household with more accurate and better information about employment opportunities. The effect of this social network was found to be positive by Mlatsheni & Rosphabé (2001), using OHS 1999 data. Dinkelman (2004) found, using KIDS data, that having employed people in the household had a negative (though not statistically significant) impact on the probability of employment.

In this case, the marginal effect of the proportion of employed adults in the household is negative and highly significant[4]. While this may seem counterintuitive, it may not be related to networks, but could point to the fact that the presence of employed adults decreases the pressure on the young people in the household to work. Of course there may be many other unobservable characteristics that are influencing the size and sign of this marginal effect, and there could possibly be some omitted variable bias present.

The gender and education characteristics of the head of the household are also included as measures of the socio-economic status if the household. This measure is, of course imperfect in that there is no fixed definition of the household head and households are free to nominate anyone within the household as the head. This may be the oldest person, the oldest male, the person with the highest income or simply the person who was present at the time of the survey. It is not even clear that the characteristics of the head of the household are representative of the rest of the members of that household. A further discussion on these issues is however, beyond the scope of this paper. If the head of the household is female, this impacts negatively on the probability of being employed. This is not surprising as female headed households are usually worse off than male headed households. The effect the education of the household head is negative, which is not what one would expect. However, if the household head has a post matric qualification, this impacts positively on the young adults' probability of being employed. That said, none of these effects are statistically different from zero.

Two proxies for the composition of the household are also included. The presence of pre-school children and elderly people could result in young people being forced to stay at home and take care of these people. Alternatively, the social grants that a household receives on account of having children and elderly people[5], add to the

[4] The regression was also run using a dummy for the presence of an employed adult the household, and again with the number of working adults in the household. The results were found to be consistent across all three specifications.

[5] Child Support Grant (R190 per month) and State Old Age Pension (R820 per month) respectively

overall income of the household, thus allowing for more resources to facilitate search activities. In this case, even though the marginal effects of the two proxies are both significant, they have different signs. The presence of pre-school children in the household increases the probability of being employed, while the presence of elderly people decreases it. It is difficult to establish which factors are driving these results, but it is clear that the composition of the household has a significant effect on young people's labour market outcomes.

Schooling characteristics

In the specification in column 4, the results of the Literacy and Numeracy Evaluation (LNE) are introduced. These scores are included as proxies for quality of education, using an output based measure of quality. Contrary to a priori expectations, the LNE scores have no impact on the probability of employment. The marginal effects are very small and both are not statistically significant. The model in column 5 includes proxies for quality of education using an input based measure of quality. We include the pupil-teacher ratio, and find a positive, significant effect. This indicates that young people who went to schools with higher teacher pupil ratios are more likely to be employed. This result is counter-intuitive, as one would expect a higher pupil-teacher ratio, which would suggest a lower quality of education, to have a negative impact on employment.

The second variable represents the proportion of privately paid teachers compared to the total number of teachers. State schools have the option of employing additional teachers, over and above the teachers that are paid by the state. These teachers are selected and paid by the governing body of the school. Only schools that are relatively better off would be able to afford additional teachers, and one would expect such schools to also be better resourced in other areas. The proportion of privately paid teacher therefore gives an indication of the quality of the school. The effect of this variable, while positive, is not significant. The effect of quality of education on the probability of employment is therefore questionable. One needs to bear in mind however, that children who receive better quality education are likely to persist in their schooling, and would thus be under-represented in this sample. Furthermore, the people in our sample are likely to be looking for low-skills jobs, where it is not clear that having better quality education will have a significant payoff.

7. Transitions and results of the dynamic (panel) analysis

Table 7 showed that the employment rates didn't vary much across the two waves of the survey. In 2002, the employment rate was roughly 40% and by 2004, it had increased to 45%. If one observes two cross samples of data, one could be tempted to conclude that by 2004, an additional 5% of the sample had found employment, over and above the 40% that were employed in 2002. Even though the coefficients from the cross-section regressions look similar for the two years, it is not clear that it is the same people who were employed in 2004, but given cross-sectional data, we cannot conclude otherwise. The major benefit of panel data is that it allows one to address this very issue of labour market dynamics. Using the two waves of CAPS, we can see who moved into and out of employment, and whose labour market status remained the same. Table 7 shows the frequency and percentage of each of four labour market states in both 2002 and 2004. Searching refers to those who had engaged in active job searching in the month prior to the survey. The discouraged are defined as those who would like a job, even though they were not searching.

Table 7: Employment status in 2002 and 2004 – Frequency and percentages

Labour Market Status	2002		2004	
	Frequency	Percent	Frequency	Percent
Not in Labour Force	106	9.40	229	20.38
Searching	283	25.19	215	19.18
Discouraged	284	25.28	168	14.95
Employed	451	40.12	511	45.49
Total	**1,123**	**100**	**1,123**	**100**

Source: CAPS 2002 and CAPS 2004

Notes:

1. All data have been calculated using sample weights.
2. Searching indicates those that have searched for work in the month prior to the survey

Over and above the difference in the employment rate, there is a big difference in the number of people who are not in the labour force. There are almost twice as many people who are not in the labour force in 2004 compared to 2002. There are more discouraged workers in 2002 than in 2004, with 25.28% and 14.95% respectively in each of the two years.

These transition rates for the sub-sample that appeared in both 2002 and 2004 are captured in table 8. Transition rates are calculated by taking the number of observations who were in state i in 2002 and state j in 2004 as a percentage of the total number of observations in state i in 2004. The entries along the diagonal indicate observations that remained in the same labour market status in 2002 and 2004.

Table 8: Transition rates from 2002 to 2004

2002 Labour Market Status	2004 Labour Market status			
	2004 Not in Labour Force	2004 Searching	2004 Discouraged	2004 Employed
2002 Not in Labour Force	**41.81**	21.23	23.71	13.26
2002 Searching	22.84	**23.46**	16.18	37.52
2002 Discouraged	26.46	21.68	**24.37**	27.5
2002 Employed	9.97	14.45	6.19	**69.39**

Source: CAPS 2002 and CAPS 2004
Notes:
1 All data have been calculated using sample weights.
2 Searching indicates those that have searched for work in the month prior to the survey.

Only 69.39% of those who were employed in 2002 are still employed in 2004. This means that just over two-thirds of those who were employed in 2002 were able to remain employed over the subsequent two years. Given the low skills base of the sample under observation, this result indicates that this group of youth does not enjoy a high degree of job security. This could be because they would probably be concentrated in casual or contract employment, or because low skilled labour is, unfortunately very dispensable. The corresponding figure in Dinkelman (2004), where she used KIDS data, is around 70%. Even though the two surveys (CAPS and KIDS) capture two different groups at different points in time, we do se some consistency in the job retention rate for different groups in South Africa.

Of those that were employed in 2002, 9.97% moved out of the labour force, while only 13.26% of those that were not in the labour force in 2002 had found employment in 2004. Those who were discouraged in 2002 seem to be evenly spread across the 4 categories in 2004. The biggest movers are those who were searching in 2002. Nearly 38% of them were employed by 2004, indicating that there is a payoff to searching. This is confirmed by the fact only 27% of their discouraged counterparts found employment by 2004. Of course, one could argue that difference in the 2004 employment rates of the two groups hardly justifies the costs associated with searching. Although such a cost-benefit analysis is beyond the scope of this paper, the panel regression will give a clearer indication of whether search in 2002 has a significant impact on the probability of being employed in 2004.

Table 9 and table 10 below show the transition rates for men and women respectively. One would expect less mobility for men, especially in and out of the labour force, as they would be less likely to be constrained by responsibilities in the home. Dinkelman (2004) found that fewer men moved out of employment, while fewer women moved out of the "not economically active" category over the five year period from 1993 to 1998. The figures along the diagonals of table 9 and table 10 tell a similar story.

Table 9: Transition rates for males

2002 Labour Market Status	2004 Labour Market status			
	2004 Not in Labour Force	2004 Searching	2004 Discouraged	2004 Employed
2002 Not in Labour Force	**33.74**	34.43	15.94	15.9
2002 Searching	18.9	**23.53**	11.33	46.24
2002 Discouraged	22.14	26.97	**20.13**	30.76
2002 Employed	10.64	14.54	4.98	**69.84**

Source: CAPS 2002 and CAPS 2004
Notes:
1 All data have been calculated using sample weights.
2 Searching indicates those that have searched for work in the month prior to the survey.

Table 10: Transition rates for females

2002 Labour Market Status	2004 Labour Market status			
	2004 Not in Labour Force	2004 Searching	2004 Discouraged	2004 Employed
2002 Not in Labour Force	**45.52**	15.15	27.29	12.04
2002 Searching	27.29	**23.37**	21.66	27.67
2002 Discouraged	29.08	18.46	**26.95**	25.51
2002 Employed	8.93	14.3	8.12	**68.66**

Source: CAPS 2002 and CAPS 2004
Notes:
1 All data have been calculated using sample weights.
2 Searching indicates those that have searched for work in the month prior to the survey.

The proportion of people who remained in employment is only slightly higher for men at 69.84% compared to 68.66% for women. There is, however a large difference in the proportion who remained out of the labour force, with 45.52% of the women who were out of the labour force in 2002 remaining there in 2004, compared to only 33.74% of the males. The payoff for searching is surprisingly higher for men, as many more of the men who were searching in 2002 finding employment in 2004. 46.24% of the men move from searching in 2002 to being employed in 2004, while only 27.67% of the women manage the same feat. There are also more men who move from being discouraged in 2002 to being employed in 2004.

Having thus seen that there was a lot of movement into and out of employment and the labour market, we proceed in the next section run panel regressions over the two year period. We look at individual, household and schooling characteristics in 2002 and evaluate their impact on the probability of employment in 2004. In addition to this, we include the labour market status in 2002 on the right hand side of the equation. This will give us a more nuanced idea and comparison of the impact of the 2002 labour market status on the probability of employment in 2004.

Results of Panel regressions

The results from the panel probit regressions of 2004 outcomes given 2002 characteristics are shown in table 11. The sub-sample of interest here is the group of young adults who have at most 12 years of schooling and were not enrolled in

school for the entire period from 2002 to 2004. The results are again represented as the marginal effects, evaluated at the mean for continuous variables, and evaluated as the discrete change from 0 to 1 for dummy variables. Z-statistics are presented in brackets.

Individual characteristics

The first column of table 11 shows the results of the model including individual variables only. As one would expect, both Coloured and White youth have a higher probability of being employed than their African counterparts. It is interesting to note however, that the marginal effect of being Coloured is higher than that of being White. This is likely to be a result of the addition of labour market states into the model. This means that compared to Africans and given the individual's labour market status in 2002, Coloureds are more likely to be employed than Whites. This result is consistent across all the specifications of the model.

The marginal effects of education are bigger in the panel than in the cross-section analysis and significant across all specifications. It comes as no surprise that education continues to be important in determining labour market outcomes, even for people who have been out of school for at least two years. Although it is not significant in the first specification of the model, the marginal effect of having a matric qualification is quite large across all the specifications. Having a matric qualification therefore adds more to the probability of being employed as one remains in the labour market over time. This additional impact of having a matric, over and above the effect of years of education, indicates that a matric qualification is an important signaling tool in the labour market.

The 2002 labour market states seem to reduce a lot of the explanatory power of other variables in the model. This is consistent across all four specifications of the model. Those who were discouraged in 2002 are not statistically different from the base group, those who were out of the labour force in 2002. This means that simply wanting a job and not actively seeking one does nothing to change an individual's labour market status two years later. Searching, on the other hand is a significant determinant of employment, albeit at the 10% level. What seems to have a big impact on the probability of employment in 2004 is being employed in 2002. Those who are employed in 2002 are up to 45% more likely to be employed than those who are not in the labour force. This could indicate that those young adults who manage to find employment tend to retain those jobs over the period under review. It could also mean that it is easier to find a new job if one has previous working experience.

Household characteristics

The two household factors that were the most important determinants of employment in the cross-section analysis in section 6 were per capita income and the number of employed adults in the household. Neither of these is significant in the panel regression. The labour market states however, continue to be a big determinant of employment. This result suggests that over time, household income and the employment states of the adults in the household are not as important as the young adult's own labour market experiences. It also suggests that relying on

other employed member of the household for labour market information is not effective in finding employment if one is not actively searching or has been employed before.

It is interesting to find that controlling for all other factors, including the labour market states in 2002, youth from female headed households are almost 9% more likely to be employed. The literature usually suggests that female headed households are usually worse off than male headed households when it comes to socio-economic and labour market outcomes. Our result seems therefore to be counter-intuitive as one would expect the marginal effect to be negative. The marginal effects of having pre-school children and elderly people in the household are smaller than was the case in the cross-section regressions. The effect of having children in the household is significant in the second specification, but not when the schooling variables are included. The effect of living in a household with elderly people is not significant.

Schooling characteristics

Unlike the case in the cross-section regressions, the pupil-teacher ratio is no longer significant. The sign of the coefficient is, however, in line with the a priory expectation that a higher pupil-teacher ratio should result in a decreased probability of employment. The LNE scores remain unimportant in explaining employment. These results suggest that quality of education may be important when one enters the labour market for the first time but over time, they become less important as one's experiences in the labour market.

The overall picture that is emerging from this panel analysis is that over time, household characteristics and quality of schooling become less important determinants of employment. The biggest contributors are gender, race, education and previous labour market status. People who were employed in 2002 are more likely to be employed two years later. There is also some payoff to searching. It is noteworthy, though not surprising that race and gender continue to play a significant role in determining labour market outcomes, even after controlling for all the other factors under consideration, particularly labour market status in 2002. Although one cannot automatically conclude that this points to labour market discrimination, the differences across racial and gender lines are still quite stark.

Table 11: Determinants of employment among youth – Panel probit regression results for 2004 outcomes and 2002 characteristics (Marginal effects)

	(1) Individual Variables	(2) Household variables	(3) Schooling (LNEs)	(4) Schooling (SRN)
Individual Variables				
Female	-0.143	-0.164	-0.137	-0.162
	(3.01)**	(3.90)**	(3.65)**	(3.88)**
Coloured	0.401	0.413	0.401	0.409
	(7.19)**	(9.26)**	(11.02)**	(9.81)**
White	0.335	0.303	0.327	0.318
	(2.08)*	(2.20)*	(3.04)**	(2.36)*
Education	0.041	0.04	0.034	0.035
	(2.32)*	(2.54)*	(2.54)*	(2.45)*
Matric 2002	0.112	0.125	0.158	0.152
	(1.46)	(1.93)+	(2.63)**	(2.38)*
Age 2002	-0.003	0.006	0.005	0.007
	(0.21)	(0.52)	(0.47)	(0.65)
Searching in 2002	0.152	0.126	0.148	0.133
	(1.6)+	(1.54)+	(1.86)+	(1.65)+
Discouraged in 2002	0.107	0.091	0.107	0.083
	(1.12)	(1.09)	(1.34)	(1.02)
Working in 2002	0.453	0.384	0.39	0.393
	(5.03)**	(4.92)**	(5.12)**	(5.10)**
Household Variables				
Log of HH Income		0.047	0.007	0.019
		(1.27)	(0.23)	(0.61)
Female Head		0.089	0.079	0.084
		(1.91)+	(1.95)+	(2.04)*
Education of Head		-0.005	0.000	0.004
		(0.55)	(0.04)	(0.47)
Head with Tertiary qualification		-0.061	-0.012	-0.036
		(0.57)	(0.12)	(0.38)
Nr Employed		-0.029	-0.009	-0.014
		(1.11)	(0.35)	(0.58)
Nr of pre-school children		0.063	0.032	0.038
		(2.14)*	(1.1)	(1.28)
Nr of elderly people (age>60)		-0.052	-0.02	-0.019
		(1.18)	(0.54)	(0.49)
Schooling variables				
Pupil-Teacher Ratio			-0.007	
			(0.97)	
Private Teachers			-0.318	
			(0.72)	
LNE Literacy Score				-0.001
				(0.2)
LNE Numeracy score				0.003
				(0.49)
Observations	*1122*	*994*	*741*	*977*

Source CAPS 2002 and CAPS 2004
+ significant at 10%; * significant at 5%; ** significant at 1%
Notes:

1. Robust z-statistic in brackets
2. For 2002 labour market status, the reference group is those who are not in the labour force

8. Conclusions

The high and growing unemployment level in South Africa cannot possibly be sustainable. More and more people enter the labour market each year and a large proportion of those are school-leaving youth. This results in young people being disproportionately affected by unemployment. In our sample, only 41% of young people who were no longer in school were employed. The shift in the economy from primary to secondary and tertiary sectors has resulted in skills biased growth and an increase in the demand for highly skilled labour at the expense of those at the lower end of the skills spectrum. Those who are competing for low skills jobs are therefore particularly vulnerable.

Twelve years after the democratization of South Africa, the importance of race and gender in determining employment has not diminished. The results from both the cross section and the dynamic analyses confirm that race and gender are important determinants of employment in Cape Town. There are many factors that feed into this result, and one is best advised to be cautious in interpreting these results as labour market discrimination.

Household income is cited in the literature as being important for making available resources that can facilitate search activities. The benefits of having other employed people in the household have also been emphasized, particularly with regard to access to labour market information. The findings from the cross-section analysis confirm the importance of these factors. In the panel regression however, where we control for search and employment in the previous period, we find that searching or being employed in 2002 is an important determinant of employment two years later. Household income and the presence of employed adults in the household are no longer significant. Therefore, given the fact that one was already searching or employed, household income and the employment status of others in the household are not important in determining employment two years down the line. Passive methods of search, which are employed by discouraged workers who probably rely on information from others regarding job opportunities, are not effective.

The results regarding quantity of schooling are in line with human capital theories that suggest that more education should result in better labour market outcomes. We find that years of education and in particular a matric qualification, are really important in the low skills environment that our sample would typically find themselves in. Quality of education however, has not proven to be as important as expected. This however, does not mean that quality is not important. One needs to bear in mind that quality of education is notoriously difficult to measure, and our measures, while in line with those often used in the literature, may not necessarily be the most precise way of capturing quality. Further research into this area would definitely shed more light on the issue.

While the LNE scores were not significant in this analysis, upcoming work by Lam and Leibbrandt suggests that the biggest impact of these LNE scores is in determining progress through school. It could be the case that the people in our sample, most of whom failed to complete their high school education, left school because their progress was hindered by poor numerical and language skills. The impact of the LNEs was most significant at that stage, and not in determining employment once one has exited the schooling system.

The most vivid story that comes from the panel analysis is the importance of previous labour market states in determining employment. The results suggest that labour market states are to a certain extent entrenched over time, particularly for those who are employed and those who are discouraged or out of the labour force. Those who were previously employed remain employed or find it easier to secure employment later. There is a payoff to searching as those who are actively searching are more likely than their passive counterparts to be employed in subsequent periods. Household characteristics lose their importance in the dynamic setting. While household socio-economic conditions may establish whether or not one is able to search for employment, once that is determined, they cease to be significant determinants of employment.

The research into youth unemployment has not been exhausted. The issues around race and gender inequalities need further exploration, as do those around quality of education. There could also be more work done to compare those with high school qualifications to others who have continued on to further education. This would give a broader picture of the entire labour market in which young people are operating and competing for employment. This will be possible as more waves of the CAPS become available over time. The literature on youth unemployment would also benefit from a national panel study of youth and their labour market experiences. This study focuses on the Cape Town area, but the problem of youth unemployment is a national one, and needs to be addressed as such.

Bibliography

Anderson, KG & Lam, D (2003) "DYNAMICS OF FAMILY STRUCTURE AND PROGRESS THROUGH SCHOOL IN SOUTH AFRICA: EVIDENCE FROM RETROSPECTIVE HISTORIES" Population Association of America, Minneapolis, Minnesota, May 2003

Archer, S (2003), "EDUCATION AND ECONOMIC DEVELOPMENT: RESOURCES, INSTITUTIONS AND TRAINING IN THE SEARCH FOR POLICY GUIDELINES IN SOUTH AFRICA". Cape Town: UCT School of Economics (mimeo), pp 68-87

Ashton, D (2005), "HIGH SKILLS: THE CONCEPT AND ITS APPLICATION TO SOUTH AFRICA". Journal of Education and Work 18, pp 19-32

Becker, G (1964) "HUMAN CAPITAL: A THEORETICAL AND EMPIRICAL ANALYSIS" National Bureau of Economic Research, New York

Bertrand, M, EF Luttmer and S Mullainathan (2000), "NETWORK EFFECTS AND WELFARE CULTURES" Quarterly Journal of Economics, 115(3), pp1019-1055, August 2000

Bhorat, H (2004) "LABOUR SUPPLY AND DEMAND CONSTRAINTS ON EMPLOYMENT CREATION: A MICROECONOMIC ANALYSIS" Unpublished, University of Cape Town

Case, A & A. Deaton (1999) "SCHOOL INPUTS AND EDUCATIONAL OUTCOMES IN SOUTH AFRICA" The Quarterly Journal of Economics, 114(3), pp 1047-1084, August 1999

Chamberlain, D & Van der Berg (2002) "EARNINGS FUNCTIONS, LABOUR MARKET DISCRIMINATION AND QUALITY OF EDUCATION IN SOUTH AFRICA" Paper to DPRU/FES Conference on *Labour Markets and Poverty in South Africa* Johannesburg, 22-24th October 2002

Cichello, P, G. Fields & M Leibbrandt (2003) "EARNINGS AND EMPLOYMENT DYNAMICS FOR AFRICANS IN POST-APARTHEID SOUTH AFRICA: A PANEL STUDY OF KWAZULU-NATAL" Development Policy Research Unit May 2003 Working Paper 03/77

Department of Labour (2005) "NATIONAL SKILLS DEVELOPMENT STRATEGY: IMPLEMENTATION REPORT, 1 APRIL 2003 TO 31 MARCH 2004". Pretoria: Department of Labour

Dinkelman, T & F. Pirouz (2002) "INDIVIDUAL, HOUSEHOLD AND REGIONAL DETERMINANTS OF LABOUR FORCE ATTACHMENT IN SOUTH AFRICA: EVIDENCE FROM THE 1997 OCTOBER HOUSEHOLD SURVEY" South African Journal of Economics 70(5), pp 865-891

Holzer, HJ (1996) "EMPLOYER SKILL NEEDS AND LABOR MARKET OUTCOMES BY RACE AND GENDER" Institute for Research on Poverty Discussion Paper no.1087-1096, Michigan State University

Kingdon, G & J. Knight (2001) "UNEMPLOYMENT IN SOUTH AFRICA: THE NATURE OF THE BEAST" Centre for the Study of African Economies, Working paper 153, Oxford University

Kingdon, G and J Knight (2000), "ARE SEARCHING AND NON-SEARCHING UNEMPLOYMENT DISTINCT STATES WHEN UNEMPLOYMENT IS HIGH? THE CASE OF SOUTH AFRICA" Unpublished, Centre for the Study of African Economies. Oxford University

Lynch, L (2005), "JOB LOSSES: BRIDGING THE RESEARCH AND POLICY DISCUSSION", Bonn: IZA Discussion paper 1518

Mincer, J (1974) "SCHOOLING EXPERIENCE AND EARNINGS" NBER, Columbia University Press

Mlatsheni, C & S Rosphabé (2002) "WHY IS YOUTH UNEMPLOYMENT SO HIGH AND UNEQUALY SPREAD IN SOUTH AFRICA>" DPRU Working Paper 02/65

Moll, PG (1996) "PRIMARY SCHOOLING, COGNITIVE SKILLS AND WAGES IN SOUTH AFRICA" Chicago, May 1996

Murray Leibbrandt & Cecil Mlatsheni "YOUTH IN SUB-SAHARAN LABOUR MARKETS" Paper for the DPRU/TIPS Conference: *"African Development and Poverty reduction: The macro-Micro linkage"* October 2004

Ryan, P (2001) "SCHOOL TO WORK TRANSITION: A CROSS NATIONAL PERSPECTIVE" Journal of Economic Literature 39(1), pp 31-92

Skills Development Act, Act 97 of 1998, Government Gazette No. 19420, November 1998. Available on http://www.info.gov.za/documents/acts/1998.htm

Statistics South Africa (2005) LABOUR FORCE SURVEY, Pretoria: Statistics South Africa

Tyler, JH, RJ Murnane & J Willet (1999) "DO COGNITIVE SKILLS OF DROPOUTS MATTER IN THE LABOUR MARKET?" NBER Working paper 7101, Cambridge, MA

Van der Berg (2001) "THE ROLE OF EDUCATION IN LABOUR EARNINGS, POVERTY AND INEQUALITY" Paper to DPRU/FES Conference on *Labour markets and poverty in South Africa* Johannesburg, 15-16th November 2001

Van der Berg, S & R Burger (2002) "EDUCATION AND SOCIO-ECONOMIC DIFFERENTIALS: A STUDY OF SCHOOL PERFORMANCE IN THE WESTERN CAPE" Paper to DPRU/FES Conference on *Labour Markets and Poverty in South Africa* Johannesburg, 22-24th October 2002

Wittenberg, M (1999) "JOB SEARCH AND HOUSEHOLD STRUCTURE IN AN ERA OF MASS UNEMPLOYMENT", Unpublished, University of the Witwatersrand

Wittenberg, M (2002) "JOB SEARCH IN SOUTH AFRICA: A NON-PARAMETRIC ANALYSIS" South African Journal of Economics 70(8), pp 1163-1197

Wolf, A (2004), "EDUCATION AND ECONOMIC PERFORMANCE: SIMPLISTIC POLICIES AND THEIR POLICY CONSEQUENCES", Oxford Review of Economic Policy 20, pp 315 – 333.

Log Files for Stata Analysis

```
/*merging caps wave 1 with the SRN data*/

#delimit ;

sort y02c11sch_code;
merge y02c11sch_code using "C:\Documents and Settings\Mamello\My
Documents\Latest CAPS\schdata\y02c11.srn2000.public.dta";
tab _merge;



/* combining school data*/

#delimit;
/*private vs. public school*/

gen private17 = .;
replace private17 = 1 if y02c17q1_15=="Independent";
replace private17 = 0 if y02c17q1_15=="Public";

gen private11 = .;
replace private11 = 1 if y02c11q1_15=="Independent";
replace private11 = 0 if y02c11q1_15=="Public";

gen private = private17;
replace private = private11 if private17==.;

label variable private "Private or Public School";
label define private 0 "Public" 1 "Independent";



************************************************
** Netting HH income of each YA's income per YA **
************************************************

*create var converting inc brac's to specific nr's, midpoints

gen tempinc = .
replace tempinc = 0 if y02d16==0
replace tempinc = 50 if y02d16==1
replace tempinc = 125 if y02d16==2
replace tempinc = 175 if y02d16==3
replace tempinc = 250 if y02d16==4
replace tempinc = 350 if y02d16==5
replace tempinc = 450 if y02d16==6
replace tempinc = 550 if y02d16==7
replace tempinc = 700 if y02d16==8
replace tempinc = 900 if y02d16==9
replace tempinc = 1125 if y02d16==10
replace tempinc = 1375 if y02d16==11
replace tempinc = 1625 if y02d16==12
replace tempinc = 1875 if y02d16==13
replace tempinc = 2250 if y02d16==14
replace tempinc = 2750 if y02d16==15
replace tempinc = 3500 if y02d16==16
replace tempinc = 4500 if y02d16==17
```

```
replace tempinc = 5500 if y02d16==18

label variable tempinc "Midpoints of YA income brackets"

*combining the two income vars

gen yaincome =.
replace yaincome = y02d15 if y02d15!=999999 & y02d15!=999998 &
y02worknow==1
replace yaincome = tempinc if yaincome==. & tempinc !=. &
y02worknow==1

label variable yaincome "Current income of YA"

* Netting YA income from hhinc

sort hhid

gen tothhinc = (h02pcy*h02hhsize)
label variable tothhinc "Total household income"

gen hhinc_netya = tothhinc
replace hhinc_netya = (tothhinc - yaincome) if yaincome~=.
label variable hhinc_netya "Total hh income less YA income"

gen pcy_netya = (hhinc_netya / h02hhsize) if hhinc_netya ~=.
label variable pcy_netya "HH pcy net of YA income"

gen logpcy_netya = log(pcy_netya)
label variable logpcy_netya "Log of HH PCY net of YA income"



**********************************************************************
creating a variable for observations that are valid for 2002 and 2004
**********************************************************************

gen valid04 = .
replace valid04 = 1 if insch_04==2 & educ04<=12
replace valid04 = 0 if insch_04==1

gen validboth =.
replace validboth =1 if valid02==1 & w2worknow~=.
replace validboth =0 if valid02==1 & (insch_03==1 | insch_04==1)
replace validboth =0 if valid02==1 & w2worknow==.

label variable validboth "Valid in both 2002 and 2004"
label variable valid04 "Valid in 2004"



/*generating var for total teachers and pupil-teacher ratio*/

# delimit;
gen teacher17 = y02c17q1_10_1;
replace teacher17 = teacher17 + y02c17q1_10_2;

gen teacher11 = y02c11q1_10_1;
replace teacher11 = teacher11 + y02c11q1_10_2;
```

```
gen teachers = teacher17;
replace teachers = teacher11 if teacher17==.;

label variable teachers "Total nr of teachers in the school";
label variable teacher17 "total teachers for c17";
label variable teacher11 "total teachers for c11";

gen pupil_teacher = y02c17q1_9 /teachers if teacher17 !=.;
replace pupil_teacher = y02c11q1_9/teachers if teacher17==. &
teacher11~=.;

label variable pupil_teacher "Pupil-Teacher Ratio";



**************************************************
** Creating an education level variable for 2004 **
**************************************************

gen educ04 = educ
replace educ04 = edlvl_02 if edlvl_02~=. & edlvl_02 > educ04 &
schres_02==3
replace educ04 = edlvl_03 if edlvl_03~=. & edlvl_03 > educ04 &
schres_03==3

label variable educ04 "Highest education level in 2004"



/*creating hh level var for educ of hh head*/

/* h02a10 is rel to hh head*/
/* h02a14 is highest grade */

#delimit ;

gen temp=.;
replace temp = h02a14 if h02a10==1;
replace temp =. if h02a14==99;
sort hhid;
egen head_educ = max(temp), by(hhid);
label variable head_educ "Highest grade completed by HH head";
drop temp;

/*creating hh level var=1 of HH head has post-seconary qualif*/

gen tertiary = 0;
replace tertiary = 1 if h02a16~=.;
replace tertiary = . if h02a16==98 | h02a16==99;
label variable tertiary "Post high school qualification (0/1)";

/*tertiary = indiv level var*/

gen temp = .;
replace temp = tertiary if h02a10==1;
sort hhid;
egen head_tertiary = max(temp), by(hhid);
label variable head_tertiary "1 if HH head has post high school
qualification";
```

```
drop temp;

/*creating a hh level var for employment of hh head*/

gen temp =.;
replace temp = h02worknow if h02a10==1;
sort hhid;
egen head_work = max(temp), by(hhid);
label variable head_work "Employment status of HH head";
drop temp;

/*creating a hh level var for gender of hh head*/

gen temp = .;
replace temp = h02a5 if h02a10==1;
sort hhid;
egen head_gender = max(temp), by(hhid);
label variable head_gender "Gender of HH head";
drop temp;



**********************************************************************
**
**Creating variables for employment of members of hh & hh
composition**
**********************************************************************
**

*creating a variable for nr of employed adults in hh

sort hhid
gen temp =.
replace temp = 1 if h02worknow==1 & h02worknow~=. & h02a3 >22
egen nr_employed = sum(temp), by(hhid)
label variable nr_employed "Number of employed adults in the HH"

*Creating dummy var for presence of employed person in hh

gen adult_work=.
replace adult_work = 0 if nr_employed==0
replace adult_work = 1 if nr_employed >0 & nr_employed ~=.
label variable adult_work "At least one adult working in HH"

*Creating variable for proportion of employed people to hh size

gen prop_working=.
replace prop_work = (nr_employed/h02hhsize) if h02worknow~=. &
h02hhsize~=.
label variable prop_work "Proportion of working adults to HH size"


*Creating variable for number of children in HH

sort hhid
gen temp = .
replace temp=1 if h02a3 <=6
egen nr_kids = sum(temp), by(hhid)
drop temp
label variable nr_kids "Number of children (age<6) in HH"
```

```
*Creating variable for number of old people

sort hhid
gen temp=.
replace temp =1 if h02a3 >= 60 & h02a3~=.
egen nr_old = sum(temp), by(hhid)
drop temp
label variable nr_old "Number of old (age>60) people in HH"



************************************************
**Private teachers to total teachers in schools**
************************************************

gen temp17=.
replace temp17 = (y02c17q1_10_2/teacher17) if teacher17~=. &
y02c17q1_10_2!=.

gen temp11 =.
replace temp11 = (y02c11q1_10_2/teacher11) if teacher11~=. &
y02c11q1_10_2!=.

gen priv_teacher = temp11
replace priv_teacher = temp17 if temp11==. & temp17!=.

label variable priv_teacher "Privately paid teachers to total
teachers"

drop temp11
drop temp17



**************************
** Creating David Table **
**************************

#delimit ;

** 2002 full Sample

sum age if age <=22 [aw=weightyr];
tab gender if age <=22 [aw=weightyr];
sum educ if age <=22 [aw=weightyr];
tab educ if age <=22 [aw=weightyr];
tab race if age<=22 & race ~=5 [aw=weightyr];
tab y02worknow if age <=22 [aw=weightyr];
sum h02pcy if age <=22 [aw=weightyr];
sum head_educ if age <=22 [aw=weightyr];
sum edmoth if age <=22 [aw=weightyr];
sum edfath if age <=22 [aw=weightyr];

** 2002 valid02 sample
#delimit ;

sum age if valid02 ==1 [aw=weightyr];
tab gender if valid02==1 [aw=weightyr];
sum educ if valid02==1 [aw=weightyr];
```

```
tab educ if valid02==1 [aw=weightyr];
tab race if valid02==1 [aw=weightyr];
tab y02worknow if valid02==1 [aw=weightyr];
sum h02pcy if valid02==1 [aw=weightyr];
sum head_educ if valid02 ==1 [aw=weightyr];
sum edmoth if valid02==1 [aw=weightyr];
sum edfath if valid02==1 [aw=weightyr];


** 2002 restricted sample ie 2002 validboth
#delimit ;

sum age if validboth ==1 [aw=weightyr];
tab gender if validboth==1 [aw=weightyr];
sum educ if validboth==1 [aw=weightyr];
tab educ if validboth==1 [aw=weightyr];
tab race if validboth==1 [aw=weightyr];
tab y02worknow if validboth==1 [aw=weightyr];
sum h02pcy if validboth==1 [aw=weightyr];
sum head_educ if validboth ==1 [aw=weightyr];
sum edmoth if validboth==1 [aw=weightyr];
sum edfath if validboth==1 [aw=weightyr];

** those in 2002 but not in 2004

#delimit ;

sum age if validboth ==0 [aw=weightyr];
tab gender if validboth==0 [aw=weightyr];
sum educ if validboth==0 [aw=weightyr];
tab educ if validboth==0 [aw=weightyr];
tab race if validboth==0 [aw=weightyr];
tab y02worknow if validboth==0 [aw=weightyr];
sum h02pcy if validboth==0 [aw=weightyr];
sum head_educ if validboth ==0 [aw=weightyr];
sum edmoth if validboth==0 [aw=weightyr];
sum edfath if validboth==0 [aw=weightyr];

** 2004 Valid04 sample
#delimit ;

sum w2b_a12 if valid04==1 [aw=weightyr];
tab  w2b_a3 if valid04==1 [aw=weightyr];
sum educ04 if valid04==1 [aw=weightyr];
tab educ04 if valid04==1 [aw=weightyr];
tab  w2b_a4 if valid04==1 [aw=weightyr];
tab w2worknow if valid04==1 [aw=weightyr];


** 2004 sample that's also in 2002 ie 2004 validboth
#delimit ;

sum w2b_a12 if validboth==1 [aw=weightyr];
tab  w2b_a3 if validboth==1 [aw=weightyr];
sum educ04 if validboth==1 [aw=weightyr];
tab educ04 if validboth==1 [aw=weightyr];
tab  w2b_a4 if validboth==1 [aw=weightyr];
tab w2worknow if validboth==1 [aw=weightyr];
```

```
*****************************************************
** Transistion Matrix - variables and actual matrix **
*****************************************************

*Create variable for LM status in 2002

gen lmstatus02 =.
replace lmstatus02 = 1 if validboth==1
replace lmstatus02 = 2 if y02d44==1 | y02d45==1
replace lmstatus02 = 3 if y02d48b==1 & (y02d44==2 | y02d45==2)
replace lmstatus02 = 4 if y02worknow==1
label variable lmstatus02 "LM status in 2002"

*Create variable for LM status in 2004

gen lmstatus04 =.
replace lmstatus04 = 1 if validboth==1
replace lmstatus04 = 2 if w2lookwrk7==1 | w2b_h9~=.
replace lmstatus04 = 3 if w2wntwrknlook==1
replace lmstatus04 = 4 if w2worknow==1
label variable lmstatus04 "LM status in 2004"

*Labeling the variables

label define lmstatuslabel 1 "Not in Labour Force" 2 "Searching" 3
"Discouraged" 4 "Employed"
label values lmstatus02 lmstatuslabel
label values lmstatus04 lmstatuslabel

*the matrix

tab lmstatus02 lmstatus04 if validboth==1 [aw=weightyr], row
tab lmstatus02 lmstatus04 if validboth==1 & female==1 [aw=weightyr],
row
tab lmstatus02 lmstatus04 if validboth==1 & female==0 [aw=weightyr],
row



/* descriptive statistics*/

#delimit ;
tab race y02worknow if valid==1 [aw = weightyr], row;
tab gender y02worknow if valid==1 [aw=weightyr], row;
tab race if valid==1 [aw=weightyr], sum(educ);
tab gender if valid==1 [aw=weightyr], sum(educ);
tab race still_school if educ<=12 [aw=weightyr], row;
tab gender still_school if educ<=12 [aw=weightyr], row;

/*Household characteristics*/
tab race if valid==1 [aw=weightyr], sum(head_educ);
tab race if valid==1 [aw=weightyr], sum(head_work);
tab race if valid==1 [aw=weightyr], sum(head_tertiary);
twoway (kdensity loghhinc if valid==1 & race==1)
 (kdensity loghhinc if race==2 & valid==1)
 (kdensity loghhinc if race==4 & valid==1), ytitle(Density)
  xtitle(Log of Household per capita income)
  title(Per Capita Household Income by population group)
  subtitle(2002) note(Source: CAPS 2002)
```

```
  legend(order(1 "African" 2 "Coloured" 3 "White"));

tab race if valid==1 [aweight=weightyr], sum(h02pcy);

/* Schooling characteristics*/

tab race if valid==1 [aw=weightyr], sum(sch_prob);
tab race if valid==1 [aw=weightyr], sum(pupil_teacher);
tab race private if valid==1 [aw=weightyr], row;

twoway (kdensity lnetotalperc if valid==1 & race==1, clpat(solid))
 (kdensity lnetotalperc if race==2 & valid==1, clpat(dash))
 (kdensity lnetotalperc if race==4 & valid==1, clpat(tight_dot)),
  ytitle(Density of LNE Total) xtitle(Total LNE scores (%))
  title(Total LNE scores by population group)
  note(Source: CAPS 2002) legend(order(1 "African" 2 "Coloured" 3
"White"));

tab race if valid==1 [aw=weightyr], sum(lnelitperc);
tab race if valid==1 [aw=weightyr], sum(lnenumperc);

/*Spikeplot of years of education*/

spikeplot educ if valid==1 [aweight=weightyr],
 frac ytitle(Frequency) xtitle(Years of education)
 title(Years of education) caption(Source: CAPS 2002);



*regressions

xi: dprobit y02worknow female i.race educ matric02 age
[pweight=weightyr] if race !=3 & valid02==1
outreg using cross2, 10pct nolabel

xi: dprobit work04 female i.race educ04 matric04 age04
[pweight=weightyr] if race !=3 & valid04==1
outreg using cross2, 10pct nolabel append

xi: dprobit y02worknow female i.race educ matric02 age
logpcy_netya female_head head_educ head_tertiary
prop_working nr_kids nr_old [pweight=weightyr] if race
!=3 & valid02==1
outreg using cross2, 10pct nolabel append

xi: dprobit y02worknow female i.race educ matric02 age
logpcy_netya female_head head_educ head_tertiary
prop_working nr_kids nr_old  lnelit lnenum
[pweight=weightyr] if race !=3 & valid02==1
outreg using cross2, 10pct nolabel append

xi: dprobit y02worknow female i.race educ matric02 age
logpcy_netya female_head head_educ head_tertiary
prop_working nr_kids nr_old  pupil_teacher priv_teacher
[pweight=weightyr] if race!=3 & valid02==1
```

```
outreg using cross2, 10pct nolabel append

xi: dprobit work04 female i.race educ matric02 age
[pweight=weightyr] if race !=3 & valid02==1
outreg using panel2, 10pct nolabel

xi: dprobit work04 female i.race educ matric02 age
logpcy_netya female_head head_educ head_tertiary
prop_working nr_kids nr_old [pweight=weightyr] if race
!=3 & valid02==1
outreg using panel2, 10pct nolabel append

xi: dprobit work04 female i.race educ matric02 age
logpcy_netya female_head head_educ head_tertiary
prop_working nr_kids nr_old lnelit lnenum
[pweight=weightyr] if race !=3 & valid02==1
outreg using panel2, 10pct nolabel append

xi: dprobit work04 female i.race educ matric02 age
logpcy_netya female_head head_educ head_tertiary
prop_working nr_kids nr_old pupil_teacher priv_teacher
[pweight=weightyr] if race !=3 & valid02==1
outreg using panel2, 10pct nolabel append
```