



Conference 2006



Progress through School in Urban South Africa: Evidence from Panel Data

David Lam, Cally Ardington and Murray Leibbrandt

**Accelerated and Shared Growth in South Africa:
Determinants, Constraints and Opportunities**

18 - 20 October 2006

**The Birchwood Hotel and Conference Centre
Johannesburg, South Africa**



Conference organised with
support from the EU



Preliminary draft

Progress through School in Urban South Africa: Evidence from Panel Data

David Lam
University of Michigan
davidl@umich.edu

Cally Ardington
University of Cape Town
cally@stats.uct.ac.za

Murray Leibbrandt
University of Cape Town
mleibbra@commerce.uct.ac.za

Paper prepared for submission to the DPRU/TIPS conference
“Accelerated and Shared Growth in South Africa:
Determinants, Constraints and Opportunities”

October 2006

Date of draft: July 14, 2006

David Lam is Professor of Economics and Research Professor in the Population Studies Center at the University of Michigan. Murray Leibbrandt is Professor of Economics and Director of the Southern Africa Labour and Development Research Unit at the University of Cape Town. Cally Ardington is Lecturer in the Department of Statistics at the University of Cape Town. Support for this research was provided by the U.S. National Institute of Child Health and Human Development and the Mellon Foundation.

Abstract:

This paper uses the Cape Area Panel Study (CAPS), a new longitudinal survey of 4,800 young people in metropolitan Cape Town, to analyze the impact of individual, household, and school characteristics on progress through school. CAPS provides data on roughly 1,500 students who were enrolled in grades 8, 9, and 10 in 2002. We follow these students over the next three years, looking at whether they reached grade 11, grade 12, and completion of grade 12 by 2005. We document large differences in the probability of grade advancement between white, coloured, and African youth. While 86% of white 9th graders in 2002 had reached grade 12 by 2005, only 29% of Africans and 42% of coloureds had reached grade 12. Probit regressions indicate that grade advancement between 2002 and 2005 is strongly associated with the respondents' performance on the literacy and numeracy evaluation conducted by CAPS in 2002. Controlling for these test scores eliminates the white and coloured advantage over Africans in progress through school. We also estimate large effects of household income, and find a negative impact of household shocks such as the loss of employment or death of an adult household member.

Introduction

The dramatic social and political changes in South Africa over the last decade have greatly changed the opportunities and incentives facing young people. The apartheid system that ended with the 1994 election imposed restrictions on non-white South Africans in many dimensions, including what schools they could attend, where they could live, whom they could marry, and what jobs they could hold. Although opportunities have expanded, young South Africans today face many challenges. Like their parents, they will enter a labor market with high unemployment and extreme earnings inequality. Unlike their parents, they have grown up during the age of HIV/AIDS, surrounded by one of the highest HIV prevalence rates in the world.

One of the most important issues affecting the future of these youth is the quantity and quality of the education they receive. South Africa has almost universal school enrollment at the primary level, with enrollment rates remaining high into the teenage years (Anderson, Case, and Lam, 2001). The level of schooling attainment is mostly determined between age 14 and 22. These are the years when young people may drop out or fail out of secondary school, may pass or fail their grade 12 matriculation exam, and may or may not go on to post-secondary education. This paper looks at one of the most critical periods in this transition, the period that follows grades 8, 9, and 10. Using a new panel study of youth collected in Cape Town, we are able to follow 8th, 9th, and 10th graders for the next three years. We find large differences between population groups in the probability that students successfully advance three grades in school between 2002 and 2005. We analyze the determinants of these differences using a rich set of variables, including performance on a baseline literacy and numeracy evaluation, household level variables such as income and parental schooling, and indicators of school quality.

The Cape Area Panel Study

This paper takes advantage of a new household survey in metropolitan Cape Town. The Cape Area Panel Study (CAPS), a collaborative project of the University of Cape Town and the University of Michigan, is a longitudinal study of youth and their families. This paper uses the first wave of the survey, which was conducted in 2002, and the third wave, which was conducted in 2005.¹ The initial wave of the Cape Area Panel Study (CAPS) was conducted in 2002 on 4,752 young people aged 14-22, living in 3,304 households. CAPS was designed as a stratified two-stage clustered sample with stratification on the predominant population group living in each

¹ See Lam and Seekings (2005), for details about CAPS, including sample design and response rates.

sample cluster. Cape Town has three predominant population groups – coloured, African/Black, and white. The distribution of the Cape Town population the 2001 census was 48% coloured, 32% African, and 19% white, with about 2% classified as Indian or other groups. Given this distribution, CAPS oversampled areas classified as predominantly African and white in order to produce larger samples of African and white respondents than would be present in a simple random sample. Cape Town is the only major city in South Africa to have substantial numbers of white, coloured, and African residents, providing unique opportunities for the study of the changing nature of inequality after the abolition of apartheid.

Wave 1 of CAPS contains two major sources of data. First, the survey includes a household questionnaire, in which demographic data on the entire household is collected. Second, the survey includes a detailed young adult questionnaire, which collects data on schooling, employment, and fertility of household members between the ages of 14 and 22. The young adult questionnaire includes a life history calendar that provides retrospective information on schooling, living arrangements, employment, fertility, and sexual partnerships. It also includes a basic numeracy and literacy skills test administered to each youth respondent. The results of this test will be used in the analysis below. CAPS youth respondents were interviewed a second time in either 2002 or 2003, and were interviewed a third time in 2005. The Wave 1 and Wave 3 data will be the major focus of the analysis in this paper.

Table 1 shows the sample size in Wave 1 and Wave 2 for respondents who were enrolled in Grade 8, 9 or 10 in 2002. This is the group that will be the major focus of the analysis below. As seen in Table 1, there were roughly 1,500 respondents in Grades 8-10 in Wave 1, 49% of whom were African. As seen in the “weighted percent” column, when we adjust for the oversampling of African respondents the African group is 32% of those enrolled in Grades 8-10. The white sample is considerably smaller, a result of both the intentional sample design and the lower response rate among white households. As with most household surveys in South Africa, response rates in CAPS were relatively high in African and coloured areas and disappointingly low in white areas. Household level response rates were about 89% in African areas, 83% in coloured areas, and 46% in white areas, for an overall household response rate of 74%. Young adult response rates, conditional on the participation of the household, were quite high, even in white areas. Given participation of the household, response rates for the selected young adults were 93% in African areas, 88% in coloured areas, and 86% in white areas. Combining the household and young adult response rates, the proportion of potential young adult respondents who actually completed questionnaires was 83% in African areas, 72% in coloured areas, and

42% in white areas. While we think it is important to include the white results in our analysis, it is important to keep in mind that the white sample may not be representative of the population of all white young adults in Cape Town.

In this paper we compare outcomes for African, coloured, and white youths. These three population groups were subject to very different treatment under apartheid, many of which may likely to continue effect young people in the post-apartheid period. Whites had advantages in a wide range of areas, including significantly higher expenditures on schooling, privileged access to the labor market, unrestricted residential mobility, and better access to most social services. Africans had the least access to services and the most restrictions on work and migration, with a large gap in expenditures on schooling. The coloured population, which is heavily concentrated in Cape Town and the Western Cape province, occupied an intermediate status under apartheid, with higher expenditures on schooling and better access to jobs.

Levels of enrollment and grade repetition

Figure 1 shows three important indicators of schooling at each age from 6 to 20 based on the retrospective reports of the CAPS respondents who were age 20-22 in 2002. The results are broken down by gender and population group. The top panel shows the proportion of respondents who were enrolled in school or post-school educational institution at each age. There are several important features about the age profile of school enrollment. The first is that enrollment rates are high, with enrollment rates for all groups are close to or above 90% for all ages between 9 and 15. A second important feature is that female enrollment rates are slightly higher than male enrollment rates for all three population groups until around age 18. The figure shows that Africans lag behind in starting school, with similar patterns for males and females. Only about 80% of African males and females were in school at age 8, compared to 99% for coloured and white males and females. Above age 9 Africans have enrollment rates of 95% to 99%, similar to those of coloured and white youth. Another important feature of the figure is the fact that Coloured enrollment rates begin to fall above age 15, with Africans having higher enrollment rates than Coloured youth at all ages above 15. African enrollment rates are about 85% at age 17 and 70% at age 18.

The second panel of Figure 1 shows the number of grades completed at each age for our 20-22 year-old respondents. The figures shows that white males and females advance almost one grade of school per year on average, reaching a mean of about 8 grades completed by age 14. Although coloured youth start school at a similar age as whites, and have almost identical

enrollment rates, the figure shows that they lag behind white youth in grade advancement from an early age. By age 14 coloured females were about 0.5 grades behind white females, with a similar gap between white males and coloured males. Africans start school later and their age profile of grade advancement has a lower slope. By age 14 African females had completed 6.4 grades and African males had completed 5.8 grades. The gap between African males and white males was already two full grades by age 14. Because of the high enrollment rates for Africans in the late teens, African grade attainment almost catches up with coloured grade attainment by age 20.

The second panel of Figure 1 also shows a female advantage in grade attainment in all three groups. As pointed out by Anderson, Case, and Lam (2001), girls move through school faster than boys in South Africa, with female schooling exceeding male schooling by about one full grade among recent cohorts of Africans who have finished schooling.

The top two panels suggest that grade repetition may be an important factor in determining grade attainment. One of the valuable features of the CAPS data is that it provides direct measures of grade repetition. For each grade of schooling respondents were asked whether they passed the grade, failed the grade, or dropped out before completing the grade. The bottom panel of Figure 1 shows the cumulative number of grades failed at each age, as reported by our respondents age 20-22.

A Stochastic Model of Grade Repetition

In this section we develop a theoretical model that we believe provides a useful framework for understanding patterns in school enrollment and progress through school in South Africa. Figure 1 documents two important features about education in South Africa. The first is that enrollment rates are relatively high. In the case of Africans and whites these enrollment rates continue to be high into the late teenage years. The second important feature is the high rate of grade repetition, especially for Africans. Figure 1 demonstrates that the schooling gap between whites and Africans at age 18 appears to be much more the result of differences in the probability of advancing from grade to grade than it is from dropout rates or non-enrollment. Making schooling decisions in an environment with high failure rates may lead to a number of important outcomes. One is that students may develop a perception, possibly correctly, that there is a high degree of randomness in the determination of which students get passing grades. Given the imperfections present in any large public school, especially one that is crowded and underfunded, it would not be surprising if there were in fact some stochastic component in the

process determining grade advancement. Whether real or perceived, a high failure rate with a large stochastic component could create an environment with some surprising incentive effects. To model this environment, consider a simple stochastic model of grade repetition. Suppose that students are evaluated at the end of the school year based on a final score S . The score for the i th student can be described the following equation:

$$S_i = \boldsymbol{\beta}'\mathbf{X}_i + u_i, \quad (1)$$

where \mathbf{X} is a vector representing the systematic determinants of student performance and u is a stochastic component that is uncorrelated with the variables in \mathbf{X} . The variables in \mathbf{X} include variables associated with the student, such as ability, attendance, and effort, variables associated with the school and neighborhood, such as measures of school quality, and variables associated with the home environment, such as parental schooling, number of siblings, and household income. Assume that $u \sim N(0, \sigma)$. This stochastic disturbance corresponds to the common perception of students that that random “luck” plays a role in their grades. It could be thought as capturing random errors made in grading schoolwork, capriciousness in teachers’ grading that is uncorrelated with actual performance, and other random elements that affect the student’s grade. We will assume that student’s perceptions about the process driving Equation (1) are consistent with reality, but for most purposes it is only necessary to assume that Equation (1) describes the perceptions of students or their parents.

Students pass the current grade if $S_i > T$, where T is a threshold established for all students at the same grade. The probability of passing is given by

$$P(S_i > T) = P[u_i > T - \boldsymbol{\beta}'\mathbf{X}_i] = 1 - \Phi\left[\frac{T - \boldsymbol{\beta}'\mathbf{X}_i}{\sigma}\right], \quad (2)$$

where Φ is the cumulative of the standard normal distribution. For simplicity we will define as the gap between

The effect of characteristics on passing

We can use Equation (2) to analyze the impact of characteristics on the probability of passing. Consider some characteristic which is one of the components of \mathbf{X} , such as previously acquired human capital, parental education, or a measure of school quality. Denote this variable by X_l , and its corresponding coefficient in Equation (1) by β_l (we will assume that $\beta_l > 0$ to simplify the discussion). To be concrete, consider the impact of mother’s schooling on the probability of passing, assuming that one year of mother’s schooling increases a student’s score

by β_1 points. To see the marginal effect of X_I on the probability of passing, we differentiate Equation (2) to get:

$$\frac{\partial P(S_i > T)}{\partial X_{Ii}} = \frac{\beta_1}{\sigma} \phi\left(\frac{T - \beta' \mathbf{X}_i}{\sigma}\right) = \beta_1 f(T - \beta' \mathbf{X}_i), \quad (3)$$

where ϕ is the density of the standard normal distribution and f is the density of the normal distribution with mean zero and standard deviation σ . Equation (3) has a number of interesting implications. First, since the density of u is highest at the mean (assumed to be zero), the effect of an increase in X_I will be largest for students who would be close to the passing threshold independent of their draw from u . Students in either the high end or the low end of the distribution will have little effect on their probability of passing if they raise their score by one more point, while students close to the threshold will have a large effect.

It is also clear from Equation (3) that the marginal effect of characteristics depends on the standard deviation σ . Evaluated near the mean, the effect of X_I is a negative function of the standard deviation σ . For those near the passing threshold, a higher variance in the random component of the score will reduce the marginal payoff to additional effort. The intuition is straightforward. The only time an increase in the score will affect whether the student passes or fails is when the student is just below the threshold. For students who would have been at the threshold based on the deterministic component alone, their probability of being at the threshold decreases as the variance of the stochastic component increases. Defining $g_i = T - \beta' \mathbf{X}_i$ as the gap between the deterministic component of the score and the threshold, and taking the derivative of Equation (3) with respect to σ ,

$$\frac{\partial^2 P}{\partial X_{Ii} \partial \sigma} = \left(\frac{\beta_1}{\sigma}\right) f(g_i) \left[\frac{g_i^2}{\sigma^2} - 1\right]. \quad (4)$$

The cross-partial derivative in Equation (4) is negative when $g_i < \sigma$ and positive when $g_i > \sigma$. As shown in Figure 1, when the deterministic component of the student's score is within one standard deviation of the passing threshold, an increase in the variance of the stochastic component will reduce the impact of characteristics on the probability of passing. Consider two students with identical characteristics who are in schooling systems that differ only in the variance of the stochastic component, with $\sigma_2 = 2\sigma_1$, and with the same β_1 coefficient in both regimes. If both students are at the passing threshold, $\beta' \mathbf{X}_2 = \beta' \mathbf{X}_1 = T$, then the marginal effect

of ability on the probability of passing will be twice as high for the student in the low-variance environment.

When the deterministic component is in either tail outside one standard deviation of the threshold, an increase in the stochastic variance will increase the impact of characteristics. Since 68% of the draws from the stochastic component will be within one standard deviation of the mean, most draws will lie in a range in which increased stochastic variance leads to a decrease in the marginal impact of characteristics on passing. This does not necessarily mean that the majority of students are in this range, however, since that will depend on the distribution of the deterministic component. It is possible, for example, that most students are well above the passing threshold and are thus relatively unaffected by the stochastic component. This will be discussed in more detail below.

Comparing two identical students who are in one of the tails of the distribution, the student in the low-variance environment will have the *lower* impact of characteristics on the probability of passing. The intuition is fairly straightforward. A student at the top of the distribution in terms of the systematic component of the score would pass the grade with certainty if the variance of the stochastic component were zero. An increase in the stochastic variance increases the probability that her score ends up close to the passing threshold, and thus increases the impact that characteristics would have on the probability of passing.

Equation (3) also reminds us of an econometric point that will be important in our empirical analysis below. If we estimate a standard probit regression of the probability of passing on mother's schooling, the regression will give us an estimate of β/σ . If we estimate regressions for two different groups with different values of σ , we will not be able to distinguish between differences in the marginal impact of mother's education on human capital accumulation (differences in β) and differences in the variance in the human capital production function (differences in σ).

How hard do students work?

One set of variables in the human capital production function is the inputs of time and effort of the student – variables such as the amount of time spent on homework and the number of days attending school. The results derived above for the impact of characteristics on the probability of passing can be applied to the impact of effort. The results imply, for example, that an increase in the variance of the stochastic component will decrease the returns to effort for students who are within one standard deviation of the passing threshold. Effort will presumably have returns in

addition to its impact on the probability of passing, but this will nonetheless be one important component of those returns.

An important difference between effort and other characteristics is that the level of effort is endogenous. Changes in the impact of effort will presumably lead to adjustments in the amount of effort supplied. Assume that students equate the marginal return to their time spent on school to the opportunity cost of that time in other uses. This opportunity cost could be the wage in labor market work or the marginal utility of leisure time (presumed to be equal for those who work). If there is a decline in the marginal returns to time spent on school, the student can be expected to reduce that time to re-equilibrate the marginal returns across all uses of time.

Who goes to school?

Assuming that school enrollment is to some extent a voluntary decision by children and/or parents, those who choose to enroll in any given school year will be those for whom the expected benefits exceed the expected costs. As an extreme simplification, suppose that attending school in any one particular year has zero payoff if the student does not pass that grade. If school enrollment required no out-of-pocket expenses and had no opportunity cost, then every student should enroll since every student has some probability of passing the grade. Even for those with deterministic components of their final score below the threshold T , there is some probability that they will get a lucky draw from u and end up with a passing score. This might be thought of as the chance that the few things they have learned in school happen to be the ones that get asked on the final exam. As σ increases, the expected probability of passing the grade is increasingly determined by the stochastic component. Suppose there were two groups of students, a low-skilled group for whom $T > \beta'X_1$ and a high-skilled group for whom $T < \beta'X_2$. If the distribution of opportunity costs were the same for the two groups, then we would expect a higher fraction of the high-skilled group to be in school. But an interesting implication of the model is that an increase in the variance of u would tend to decrease the enrollment of high-skilled students at the same time that it increases the enrollment of low-skilled students. The reason is that the probability of high-skilled students passing goes down because of the increased chance of getting draws from the bottom of the distribution. The probability of low-skilled students passing goes up because of an increased probability of getting a draw large enough to push them over the passing threshold. An increase in the variance would therefore have the potential to diminish the difference in enrollments between low-skilled and high-skilled students, *ceteris paribus*.

While the cases discussed above are simple and somewhat unrealistic, some important realism is captured by the model. The most important is that we can characterize an environment which has the following features – relatively high fractions of students fail each grade, high fractions of the eligible population are enrolled, and high fractions of students expend relatively low effort on school. These are broadly the features observed in much of the South African educational system experienced in predominantly black schools. The somewhat puzzling combination of high failure rates and high enrollment rates can be explained by an environment with a large stochastic component to grade advancement. The model has a number of empirical implications that can be analyzed using survey data such as CAPS. First, if there is an important random component to promotion rates, then the stochastic component of single year promotion rates will be higher than the stochastic component in cumulative schooling attainment. Put another way, the explanatory power of standard family-level characteristics in a regression explaining promotion rates at any given age will be smaller than the explanatory power of the same variables for cumulative schooling attainment to that age. A related implication is that enrollment rates will tend to be less affected by family background in a system with high failure rates, assuming that these higher failure rates have a large stochastic component. Given the much lower failure rates of white schools, we should find a weaker link between enrollment and family background for non-whites. The model also implies that we may find a weaker impact on progress through school of performance-related characteristics such as ability and family background among non-whites. This is because the stochastic component of grade progression weakens the impact of ability and effort.

Interactions between School and Work

Decisions by young people about whether to stay in school and how much effort to apply to school will be affected by the opportunity cost of their time. It is important to consider the labor market environment experienced by school-aged youth. Figures 2 and 3 analyze the transition from school to work as reported retrospectively by the same respondents aged 20-22 who were used in Figure 1. The figures show the percentage in each of four school/work combinations at each age – in school and not working, in school and working, working and not in school, and neither working nor in school. Work is defined broadly here, and includes any work done during the year. This includes work during school vacations, so it is important to keep in mind that the work/school combination does not necessarily imply that work was being combined with school.

Looking at the results for females in Figure 1, we see large differences in the transitions from school to work across population groups. While being in school without working is by far

the predominant activity for all three groups at age 14, by age 17 some sharp differences have emerged. Significant proportions of white females are working during years when they are still in school, with 45% of white girls in the work and school category at age 17. African females have extremely low rates of work. The percentage of African girls who work during years when they are still in school is negligible, never exceeding 3%. Less than 10% of the African females report that they were working at age 20, compared to 58% of white females. The transition from school to work for coloured females is characterized more by a sharp transition than it is for either white or African females. Relatively small proportions of coloured females work during the years they are in school, and the proportion of coloured females enrolled in school drops below that of both Africans and whites by age 16. White females have the highest proportions working at all ages, but almost always combine this work with school until age 18.

Grade Progression between 2002 and 2005

The 2005 CAPS Wave 3 data make it possible to follow the progress of young people who were still in school in 2002. We will focus on the experience of respondents who were in Grades 8, 9, and 10 in 2002. If they remained in school and passed all grades, the 8th and 9th grade students would have reached grades 11 and 12, respectively, by the 2005 wave. Those in 10th grade in 2002 who stayed in school and passed the 10th and 11th grades would have had the opportunity to take the grade 12 matriculation exam in 2004. The matric exam is nationally standardized and is one of the most important school transitions young people can make in South Africa.

Figure 4 shows the activities in 2005 of those who were in grade 8 and grade 9 in 2002, disaggregated by population group. Looking at the top panel, 91% of whites who were in grade 8 in 2002 had advanced to either grade 11 or grade 12 by 2005. The experience of African and coloured youth is very different. Among Africans who were in grade 8 in 2002, only 37% had advanced to either grade 11 or grade 12. A higher percentage, 37%, were in grade 10, implying that they had progressed two grades in three years. About 20% of Africans who had been in grade 8 in 2002 were not enrolled in 2005, with only 3% not enrolled and working. Coloured youth who were in grade 8 in 2002 were less likely than Africans to be enrolled in 2005, but those who were enrolled were more likely to have maintained normal progress through school. About 47% were in grade 11 or grade 12, with 13% in grade 10.

Characteristics affecting progress through school

In the regressions below we will use a number of individual, household, and school characteristics to predict progress through school. In this section we provide an overview of some of these characteristics. One interesting feature of the CAPS data is the numeracy and literacy evaluation that was administered to all youth respondents in Wave 1. This was a self-administered written test that was taken by the respondent after the completion of the young adult questionnaire. The test had 45 questions and took about 20 minutes to complete. The respondent could choose to take the test in either English or Afrikaans. There was no version in Xhosa, the home language of most African respondents. The English language test was taken by 99% of the African respondents, 43% of the coloured respondents, and 64% of the white respondents. In interpreting the results below it is important to keep in mind that most African respondents were taking the test in a second language, while white and coloured students were taking the test in their first language. We use the test below as a measure of cumulative learning at the time of the interview. Performance on the test reflects a combination of many factors, including innate ability, home environment, and the quantity and quality of schooling up to that point.

Figure 5 presents kernel density estimates of the distribution of numeracy and literacy test scores for each population group. Each score is standardized to zero mean unit variance. The differences in test scores across population groups are striking. Looking at the numeracy scores in the top panel, we see only a small area of overlap between the test scores of African and white respondents. The distribution of numeracy scores for coloured youth sits between, with considerable overlap with both the white and African distributions. The mean standardized numeracy score is -0.4 for Africans, 0.1 for coloureds, and 1.4 for whites, implying a two standard deviation gap between whites and Africans. This large difference in the distribution of test scores is important to keep in mind in our regressions below, where we will include the test scores as regressors.

The bottom panel of Figure 5 shows the distribution of literacy scores. The literacy test appears to have been very easy for white youth, with scores bunched at the top of the range. There was a much larger range of performance among coloured and African youth. The mean standardized literacy score is -0.4 for Africans, 0.25 for coloureds, and 0.9 for whites, with the standard deviation of African scores (1.04) more than double the standard deviation for whites (0.4). As with the numeracy scores, there is only a small range in which the African and white distributions overlap.

Another important variable in our regressions will be household income. We use the log of per capita household income at the beginning of the period, as reported by an adult respondent in the Wave 1 household questionnaire. Figure 6 plots the kernel densities for the distribution of income for each population group, standardized to the mean income for the combined population. Once again we see very large differences between population groups. The difference in mean log income between whites and Africans is about 2.4. Exponentiated, this implies that white youth in 2002 were living in homes with over 10 times higher per capita household income than Africans. As was the case with test scores, a striking feature of Figure 6 is the very small range in which the African and white income distributions overlap. The coloured distribution sits between the two distributions, overlapping more with the African distribution than with the white distribution.

An additional factor to consider in explaining school progress for 8th, 9th, and 10th graders is the extent to which students were already behind in school in 2002. As shown in Figure 1, grade repetition is an important feature of the school experience of both African and coloured youth, and by grades 8, 9, and 10 there will be considerable differences in the age of students. Figure 7 shows the distribution of ages for 8th and 9th graders in CAPS Wave 1. Looking at 8th graders in the top panel, we see that there is probably some truncation due to the fact that our sample begins at age 14. We lose some 13 year-olds who would have been in grade 8 in 2002, though this will have been a small proportion of all 8th graders. There are very large differences in the age distribution of 8th graders across population groups. White 8th graders are heavily concentrated at age 15, with less than 20% at age 16. By contrast, the modal age of African 8th graders is 15, with a wide distribution ranging between ages 14 and 21. The differences are even greater among 9th graders. About 90% of white 9th graders are age 15 or below, compared to 29% of African 9th graders and 70% of coloured 9th graders. Roughly 25% of African 9th graders are age 18 or older.

Probit regressions of progress through school

This section presents results of probit regressions in which our dependent variable is an indicator of progress through school between 2002 and 2005. We begin with a standard probit regression. The issue of heteroscedasticity, an obvious concern given our theoretical model, will be discussed below. In our first set of regressions we use the sample of 8th and 9th graders. Our dependent variable is equal to 1 if the respondent advanced at least three years in school by 2005 – 8th graders reached at least grade 11 and 9th graders reached at least grade 12. The dependent variable is equal to 0 if there is any other outcome, including dropping out of school before

reaching the target grade or being in school in some grade below the target grade in 2005. In our second set of regressions we use the sample of 10th graders in 2002. The dependent variable is equal to 1 if the respondent successfully passed the grade 12 matriculation exam by 2005, and is 0 for any other outcome. We begin with regressions that include dummy variables for population group. We then sequentially add additional regressors reflecting individual, household, and school characteristics.

Table 2 presents the probit regressions for grade advancement of 8th and 9th graders. For each regression the first column gives the regression coefficient with the robust standard error presented below it in brackets. The second column gives the marginal effect of a discrete change in the independent variable, evaluated at the sample means. Probit 1 includes a dummy for white and a dummy for coloured, with African as the omitted category. It also includes a dummy variable for grade 8, which allows the probability of advancement from grade 8 to grade 11 to have a different mean than the probability of advancement from grade 9 to grade 12. The coefficients on white and coloured are both positive and highly significant. The coefficient on white implies that white respondents had a 52 percentage point higher probability of advancing three grades than African respondents between 2002 and 2005. Coloured respondents had a 13 percentage point higher probability of advancing than African respondents. In the rest of the regressions in Table 2 we will see how these population group differences are affected by controls for individual, household, and school characteristics.

Probit 2 adds controls for three indicators of the respondent's cumulative learning in 2002. We include their literacy score, numeracy score, and the number of years they were behind in school. The "Grades behind in 2002" variable is defined as Age-Grade-6, with a lower bound of zero. We include this as an indicator of previous poor performance in school, which may also be a reflection of low school quality. The "Grades behind" variable has a negative and statistically significant effect on the probability of advancing three years, with each grade behind implying a 13 percentage point decline in the probability of advancement. The coefficient on the literacy score is positive but statistically insignificant. The coefficient on the numeracy score is negative and statistically significant at the 1% level. The magnitude of this coefficient is large, implying that a one standard deviation increase in the numeracy score increases the probability of advancing three grades by 19 percentage points. A striking result of Probit 2 is that the point estimate on the white dummy has dropped from 1.6 to 0.18 and is no longer statistically significant. The coefficient on the coloured variable has actually changed signs from positive to negative, and is statistically significant at the 1% level. Taken literally, the coefficient implies

that if we hold constant test scores and the number of grades behind in school, coloured 8th and 9th graders were 14 percentage points *less* likely to advance three grades in school between 2002 and 2005.

Probit 3 adds variables related to household resources and family background. Father's schooling and mother's schooling are measured in years of completed schooling. CAPS collected parental schooling in both the household questionnaire and directly from the youth respondents, so coresidence is not required to have information on parental schooling. In cases in which parental schooling is missing, the mother's schooling and father's schooling variables are coded as zero and the parental schooling missing variable is set to one. Mother's schooling is missing for 10% of the observations in Table 1; father's schooling is missing for 34% of the observations. Log of per capita household income is the same variable show in Figure 6 and discussed above. The household shock variable is a dummy indicating that the household respondent in 2005 reported that the household was negatively affected by an event such as the death of an adult household member, the job loss of an adult household member, or the loss of an important source of income. About 16% of respondents included in Table 1 lived in households that reported a household shock between 2002 and 2005.

Looking at the coefficients in Probit 3, per capita household income has a statistically significant positive effect on the probability of grade advancement. A 10% increase in household income would increase the probability of grade advancement by 1.05 percentage points, holding all other variables at their sample means. The 2.4 log point difference in the mean of white and African log income shown in Figure 6 would imply a $2.4 \times 0.105 = 25.1$ percentage point increase in the probability of grade advancement, holding test scores, parental schooling, and other variables in Probit 3 constant. Mother's schooling has a positive but statistically insignificant effect. The coefficient on father's schooling is larger and is significant at the 5% level, implying that a one year increase in father's schooling would cause a 2.2 percentage point increase in the probability of grade advancement. The coefficient on the household shock dummy is negative and statistically significant at the 5% level. It implies that experiencing a negative household shock between 2002 and 2005 decreased the probability of grade advancement by 14.5 percentage points.

The coefficient on both the white and coloured dummies are both negative and statistically significant in Probit 3. Taken literally, they imply that if white students had the same test scores, parental schooling, household income, and other Probit 3 characteristics as African students, the white students would have had 27 percentage point *lower* probability of advancing three grades

between 2002 and 2005. We are reluctant to take this result at face value, since we have seen above that there is very little overlap in the distributions of variables such as test scores and household income, two of the variables with very strong effects in these regressions. The regression is effectively being required to predict African and white school progress across values of the independent variables at which there are very few observations in the data. While we do not want to overstate the interpretation of these coefficients, the regressions do suggest that much of the large white advantage in grade advancement can be explained by the fact that white students had already acquired a great deal more learning than African students in grade 8 and 9. If we control for this difference, white students do not do better than African students at moving through the higher grades.

Probit 4 adds dummy variables indicating the former educational department of the school attended in 2002. Under apartheid the House of Representative (HOR) schools were intended for coloured students, and these schools continue to be predominantly coloured schools. The House of Assembly (HOA) schools were white schools, and continue to be predominantly white. HOR schools generally have considerably better infrastructure than predominantly black schools, and HOA schools are better than both black and coloured schools in most measures of school quality. The dummies are included here to pick up a wide variety of school quality. The coefficient on HOR (coloured) schools is not statistically significant, with the point estimate suggesting that students in these schools actually do worse than students in traditionally African schools. We estimate a large positive coefficient for HOA (white) schools, significant at the 5% level. The coefficient implies that holding all other variables constant, students attending white schools have a 25% higher probability of grade advancement than students attending traditionally black schools. Including these school dummies does not have a large effect on the estimated impact of test scores or household income. They do change the dummies on white and coloured, with the white coefficient becoming even more negative and the coloured coefficient becoming less negative. If we consider both the white dummy and the HOA school dummy at the same time, the results in Probit 4 imply that a white student in a white school would have a lower probability of advancing than an African student in an African school, holding all of the other variables in Probit 4 constant. As above, we are hesitant to draw too much from this surprising result, though we believe the results do point to a powerful impact of cumulative differences in learning prior to the final grades of secondary school. One potential policy implication of the results is that the large differences in secondary school grade progression between population groups cannot be eliminated simply by policies focused on these grades.

The cumulative disadvantages that African students bring with them to grade 8 and 9 have a powerful impact on their ability to advance through the final grades.

Table 3 presents a similar regressions using successful completion of the grade 12 matriculation exam as the outcome. The sample is the group of respondents who were enrolled in grade 10 in 2002, giving a sample size of about 400. The results are very similar to those in Table 2. Probit 1 shows that African 10th graders have a 48 percentage point lower probability than whites and a 16 percentage point lower probability than coloureds of successfully having passed the matric exam three years later. Probit 2 shows that controlling for numeracy scores in 2002 and the number of grades behind in 2002, these white and coloured advantages become insignificant. Being one year older (one year further behind) in 2002 decreases the probability of completing matric by 10.6 percentage points. A one standard deviation increase in the numeracy score increases the probability by 19.2 percentage points. Adding the household variables in Probit 3, only household income has a statistically significant effect. A 10 percent increase in per capita household income increases the probability of completing matric by 1.08 percentage points, very similar to the effect on grade advancement of 8th and 9th graders estimated in Table 2. Adding dummies for former school department in Probit 4, the coefficients are not statistically significant for either HOR or HOA schools. The lack of precision of estimates in Table 3 presumably reflects in part the fact that the sample is only about half as large in Table 3 as it was in Table 2.

References

- Anderson, Kermyt G., Anne Case, and David Lam (2001) "Causes and Consequences of Schooling Outcomes in South Africa: Evidence from Survey Data." *Social Dynamics* 27(1): 1-23.
- Case, Anne, and Angus Deaton (1999) "School Inputs and Education Outcomes in South Africa," *Quarterly Journal of Economics*, August, 114(3): 1047-84.
- Lam, David, and Jeremy Seekings (2005) "The Cape Area Panel Study: Technical Documentation for Wave 1," Centre for Social Science Research, University of Cape Town.
- Lloyd, Cynthia (2005) Editor, *Growing Up Global: The Changing Transitions to Adulthood in Developing Countries*, National Academies Press: Washington, DC.
- Moll, Peter (1998) "Primary Schooling, Cognitive Skills and Wages in South Africa," *Economica*, 65: 263-84.
- Thomas, Duncan (1996) "Education Across Generations in South Africa," *American Economic Review*, 86(2): 330-334.

Table 1. Sample size by population group and attrition between waves, respondents in grades 8, 9, and 10 in 2002, Cape Area Panel Study

Population Group	CAPS Wave 1, 2002			CAPS Wave 3 - 2005			Rate of Attrition
	Sample Size	Unweighted Percent	Weighted Percent	Sample Size	Unweighted Percent	Weighted Percent	
Black/African	736	48.9	32.0	578	31.99	29.84	21%
Coloured	610	40.6	51.9	549	51.91	55.41	10%
White	158	10.5	16.1	122	16.1	14.75	23%
Total	1,504	100.0	100.0	1249	100	100	17%

Table 2. Probit regressions of grade progression for 8th and 9th graders in 2002
 Dependent variable =1 if respondent advanced three grades by 2005, CAPS Wave 1 and Wave 3

Independent Variable	Probit 1		Probit 2		Probit 3		Probit 4	
Female	0.05 [0.100]	<i>0.02</i> [0.040]	0.06 [0.107]	<i>0.024</i> [0.043]	0.077 [0.110]	<i>0.031</i> [0.044]	0.081 [0.116]	<i>0.032</i> [0.046]
White	1.636 [0.206]**	<i>0.524</i> [0.038]**	0.179 [0.265]	<i>0.071</i> [0.105]	-0.724 [0.334]*	<i>-0.269</i> [0.108]*	-1.194 [0.375]**	<i>-0.403</i> [0.090]**
Coloured	0.336 [0.113]**	<i>0.133</i> [0.044]**	-0.352 [0.134]**	<i>-0.14</i> [0.053]**	-0.628 [0.158]**	<i>-0.246</i> [0.060]**	-0.401 [0.200]*	<i>-0.159</i> [0.078]*
Grade 8 in 2002	0.17 [0.103]	<i>0.068</i> [0.041]	0.297 [0.107]**	<i>0.118</i> [0.042]**	0.279 [0.109]*	<i>0.111</i> [0.043]**	0.36 [0.111]**	<i>0.143</i> [0.044]**
Grades behind in 2002			-0.399 [0.071]**	<i>-0.159</i> [0.028]**	-0.368 [0.073]**	<i>-0.147</i> [0.029]**	-0.409 [0.080]**	<i>-0.163</i> [0.032]**
Literacy score			0.167 [0.101]	<i>0.067</i> [0.040]	0.167 [0.103]	<i>0.067</i> [0.041]	0.15 [0.108]	<i>0.06</i> [0.043]
Numeracy score			0.486 [0.086]**	<i>0.193</i> [0.034]**	0.364 [0.087]**	<i>0.145</i> [0.035]**	0.333 [0.091]**	<i>0.133</i> [0.036]**
Log per capita hh income					0.264 [0.072]**	<i>0.105</i> [0.029]**	0.242 [0.073]**	<i>0.096</i> [0.029]**
Mother's schooling					0.033 [0.026]	<i>0.013</i> [0.010]	0.033 [0.027]	<i>0.013</i> [0.011]
Father's schooling					0.055 [0.023]*	<i>0.022</i> [0.009]*	0.05 [0.024]*	<i>0.02</i> [0.009]*
Mother's schooling missing					0.28 [0.298]	<i>0.111</i> [0.117]	0.304 [0.309]	<i>0.12</i> [0.121]
Father's schooling missing					0.257 [0.219]	<i>0.102</i> [0.087]	0.256 [0.224]	<i>0.102</i> [0.089]
Household shock 2002-2005					-0.373 [0.146]*	<i>-0.145</i> [0.055]**	-0.369 [0.154]*	<i>-0.144</i> [0.058]*
Former HOR school in 2002							-0.248 [0.166]	<i>-0.098</i> [0.065]
Former HOA school in 2002							0.644 [0.275]*	<i>0.251</i> [0.100]*
Constant	-0.558 [0.105]**		0.216 [0.145]		-1.88 [0.457]**		-1.777 [0.474]**	
Observations	821		814		770		731	

Robust standard errors in brackets, marginal effects evaluated at means in italics

* significant at 5%; ** significant at 1%

Table 3. Probit regressions for passing matriculation exam for 10th graders in 2002

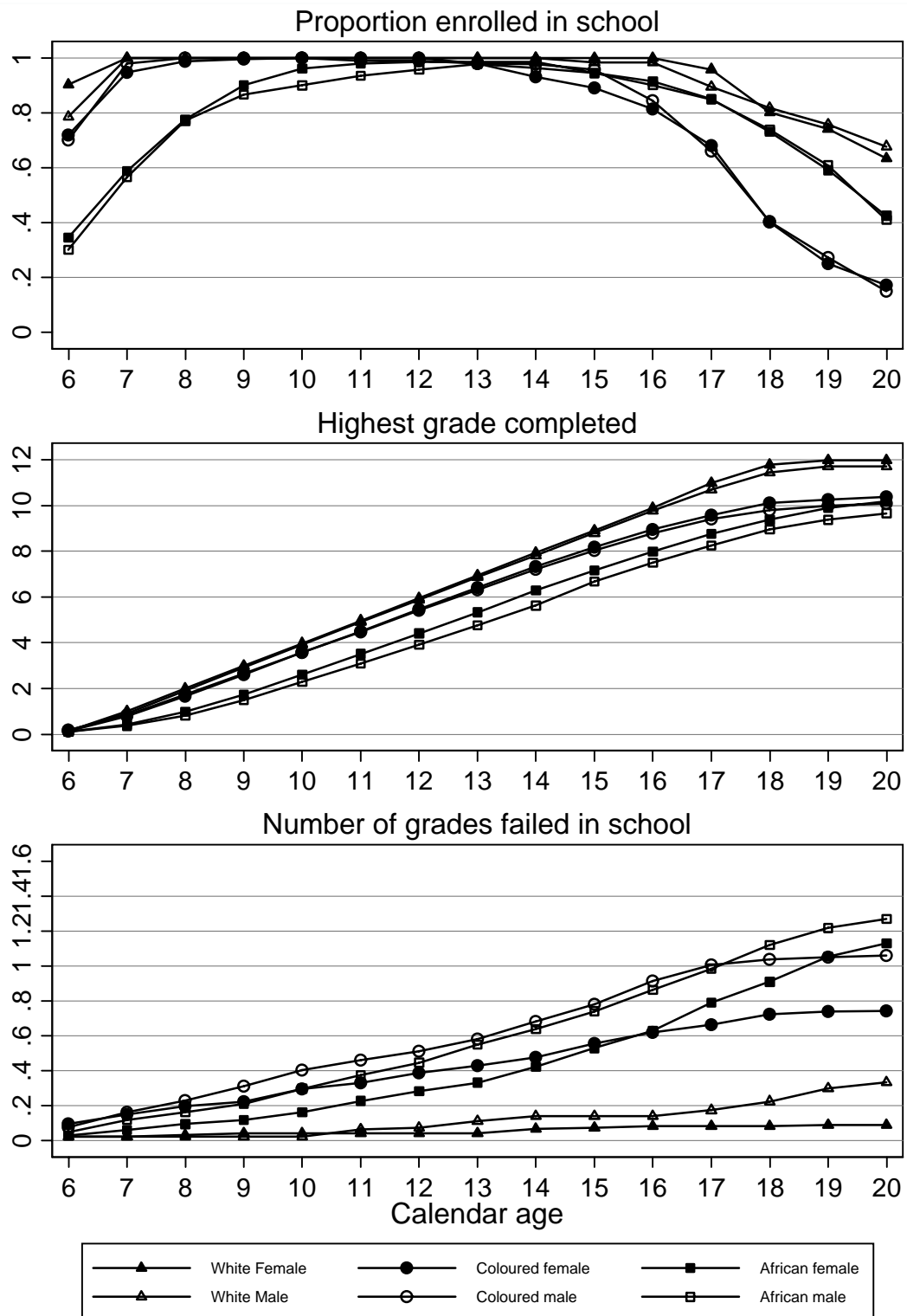
Dependent variable =1 if respondent passed Grade 12 matriculation exam by 2005

Independent Variable	Probit 1		Probit 2		Probit 3		Probit 4	
Female	0.159	<i>0.062</i>	0.241	<i>0.093</i>	0.261	<i>0.099</i>	0.166	<i>0.062</i>
	[0.138]	[<i>0.054</i>]	[0.144]	[<i>0.055</i>]	[0.148]	[<i>0.055</i>]	[0.155]	[<i>0.058</i>]
White	1.34	<i>0.48</i>	0.018	<i>0.007</i>	-0.932	<i>-0.293</i>	-1.124	<i>-0.327</i>
	[0.263]**	[<i>0.070</i> **]	[0.310]	[<i>0.120</i>]	[0.390]*	[<i>0.090</i> **]	[0.446]*	[<i>0.086</i> **]
Coloured	0.422	<i>0.163</i>	-0.147	<i>-0.057</i>	-0.475	<i>-0.181</i>	-0.321	<i>-0.122</i>
	[0.159]**	[<i>0.061</i> **]	[0.188]	[<i>0.073</i>]	[0.211]*	[<i>0.080</i>]*	[0.301]	[<i>0.114</i>]
Grades behind in 2002			-0.275	<i>-0.106</i>	-0.279	<i>-0.106</i>	-0.38	<i>-0.143</i>
			[0.081]**	[<i>0.031</i> **]	[0.086]**	[<i>0.032</i> **]	[0.097]**	[<i>0.036</i> **]
Literacy score			0.286	<i>0.11</i>	0.317	<i>0.121</i>	0.311	<i>0.117</i>
			[0.155]	[<i>0.059</i>]	[0.168]	[<i>0.063</i>]	[0.181]	[<i>0.068</i>]
Numeracy score			0.498	<i>0.192</i>	0.411	<i>0.156</i>	0.41	<i>0.155</i>
			[0.114]**	[<i>0.045</i> **]	[0.123]**	[<i>0.047</i> **]	[0.125]**	[<i>0.048</i> **]
Log per capita hh income					0.284	<i>0.108</i>	0.29	<i>0.109</i>
					[0.111]*	[<i>0.043</i>]*	[0.123]*	[<i>0.047</i>]*
Mother's schooling					0.002	<i>0.001</i>	-0.007	<i>-0.003</i>
					[0.034]	[<i>0.013</i>]	[0.036]	[<i>0.014</i>]
Father's schooling					0.031	<i>0.012</i>	0.029	<i>0.011</i>
					[0.030]	[<i>0.011</i>]	[0.031]	[<i>0.012</i>]
Mother's schooling missing					-0.243	<i>-0.089</i>	-0.498	<i>-0.171</i>
					[0.393]	[<i>0.138</i>]	[0.416]	[<i>0.126</i>]
Father's schooling missing					0.147	<i>0.056</i>	0.116	<i>0.044</i>
					[0.323]	[<i>0.125</i>]	[0.349]	[<i>0.134</i>]
Household shock 2002-2005					0.006	<i>0.002</i>	0.133	<i>0.051</i>
					[0.205]	[<i>0.078</i>]	[0.217]	[<i>0.084</i>]
Former HOR school in 2002							-0.332	<i>-0.125</i>
							[0.276]	[<i>0.103</i>]
Former HOA school in 2002							0.009	<i>0.003</i>
							[0.346]	[<i>0.131</i>]
Constant	-0.711		-0.272		-2.093		-1.827	
	[0.142]**		[0.183]		[0.695]**		[0.807]*	
Observations	426		425		399		367	

Robust standard errors in brackets, marginal effects evaluated at means in italics

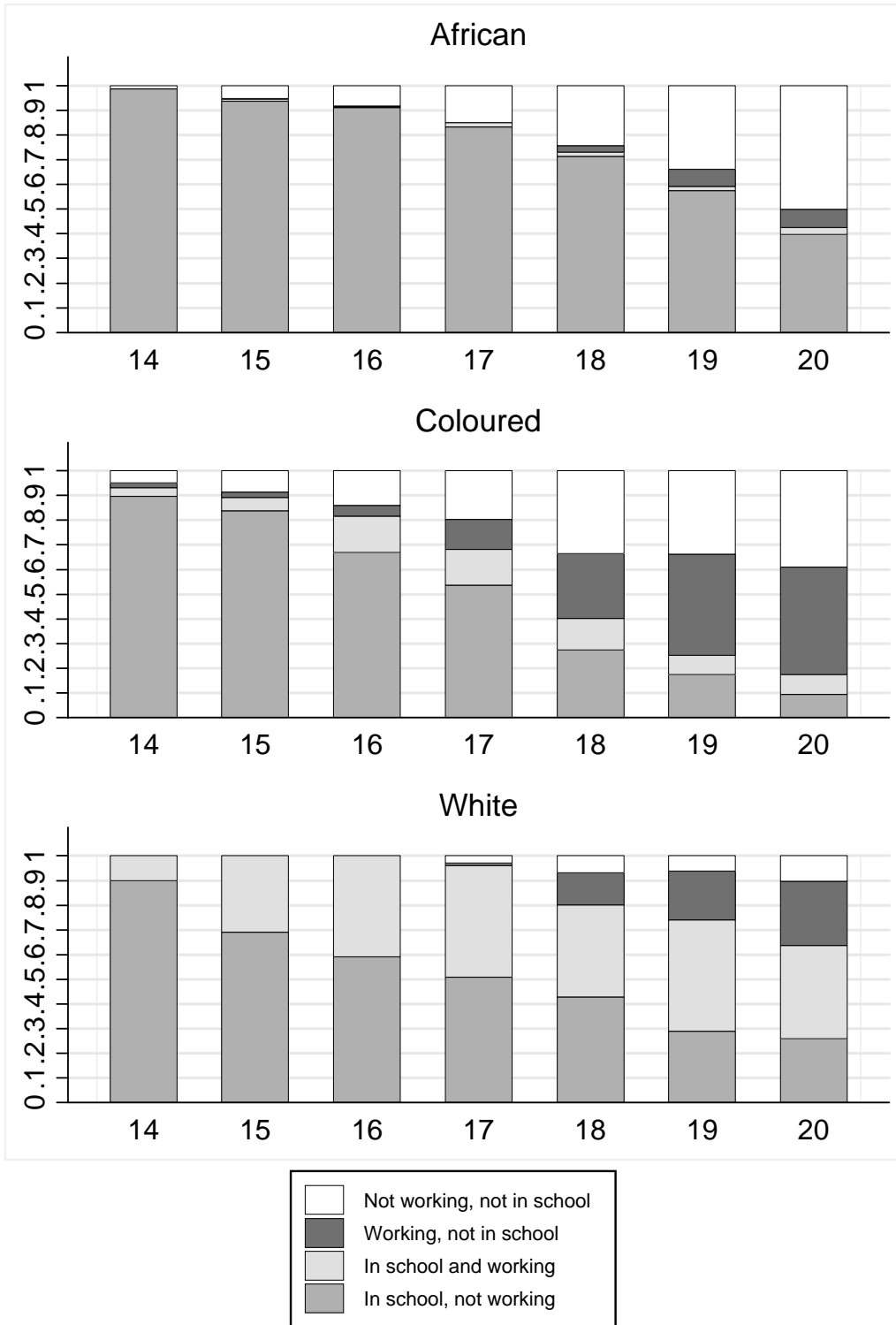
* significant at 5%; ** significant at 1%

Figure 1. Schooling experience from retrospective histories
CAPS respondents age 21-22, 2002



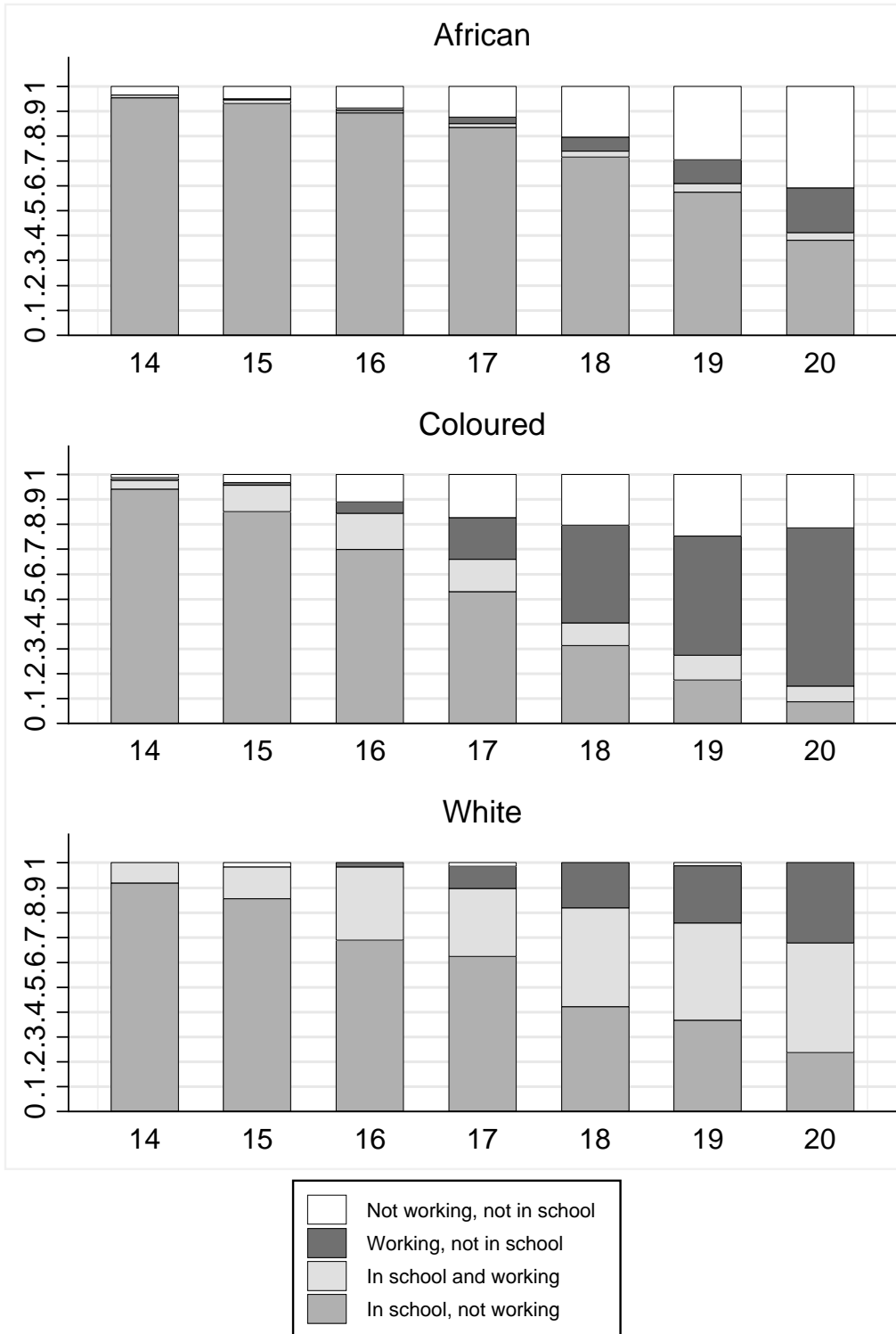
Cape Area Panel Study Wave 1, 2002

Figure 2. Categories of work and school at age 14-20
Female CAPS respondents age 21-22, 2002



Cape Area Panel Study Wave 1, 2002

Figure 3. Categories of work and school at age 14-20
Male CAPS respondents age 21-22, 2002



Cape Area Panel Study Wave 1, 2002

Figure 4. Transitions between 2002 and 2005 for 8th graders and 9th graders in 2002

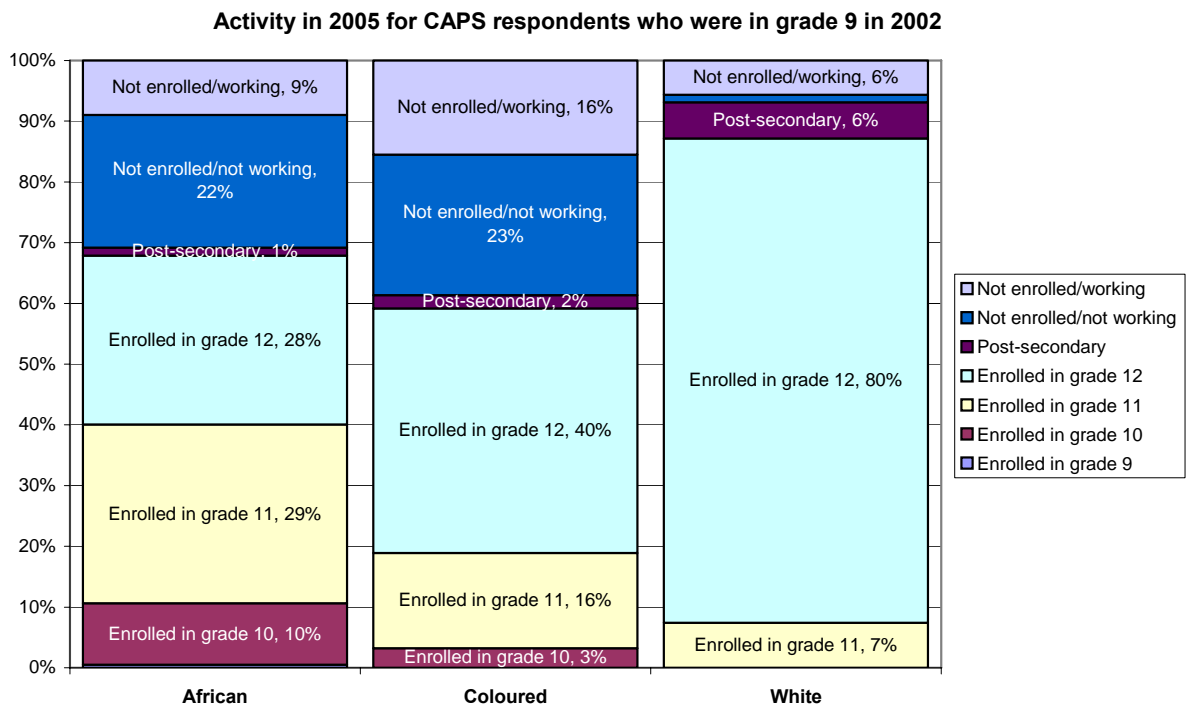
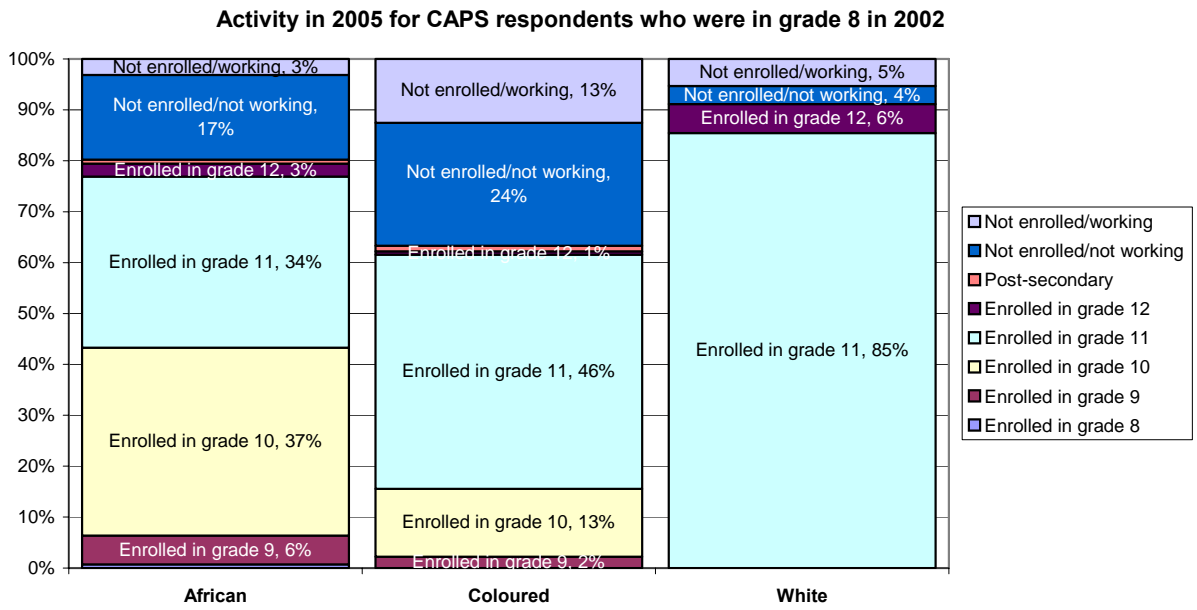
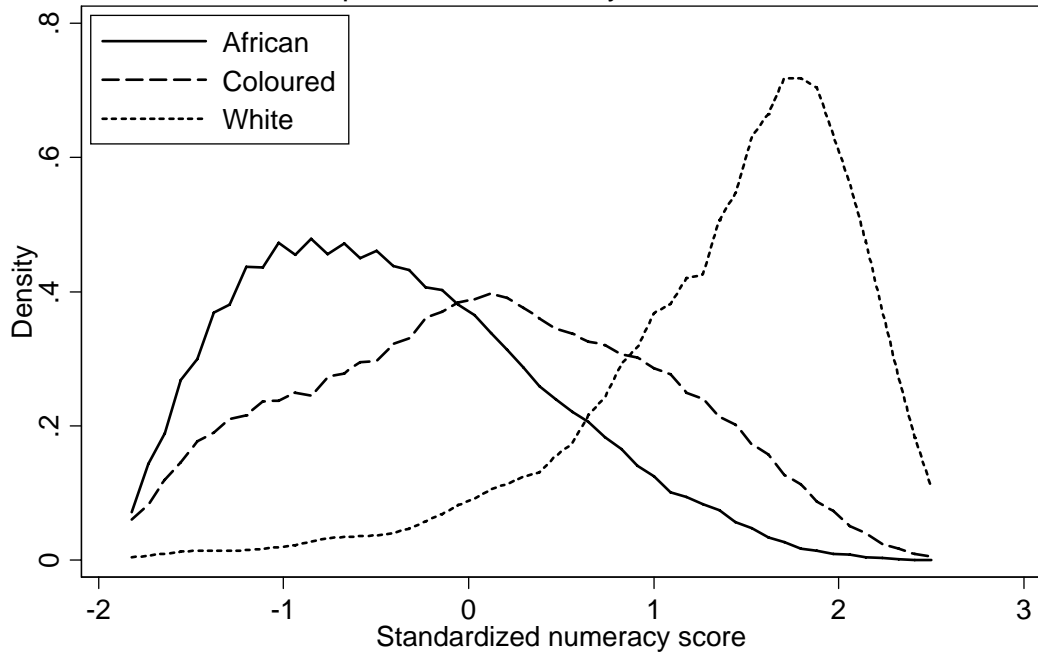


Figure 5. Kernel densities of CAPS numeracy and literacy scores

Kernel densities of standardized CAPS Numeracy scores

Cape Area Panel Study Wave 1, 2002



Kernel densities of standardized CAPS literacy scores

Cape Area Panel Study Wave 1, 2002

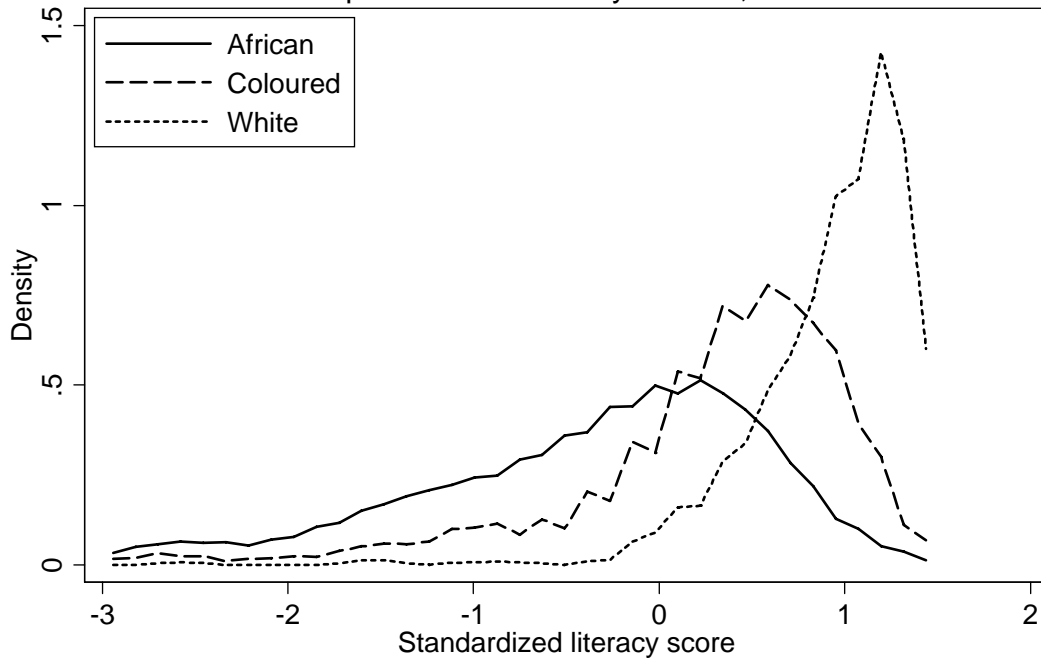


Figure 6. Kernel densities of log per capita household income, CAPS Wave 1

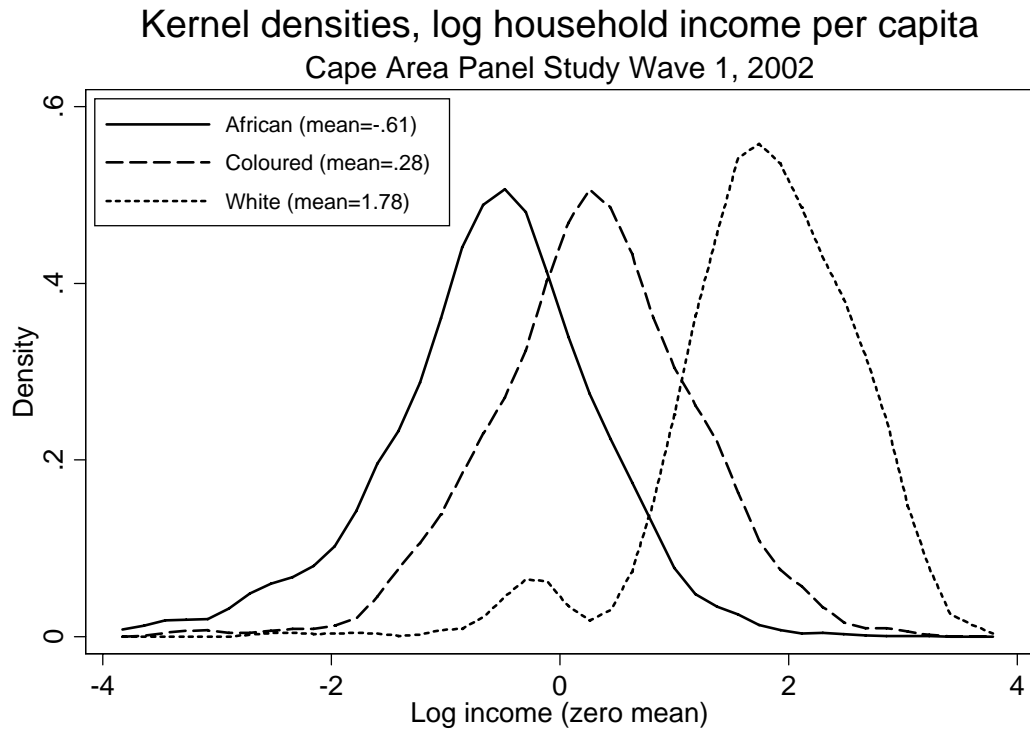


Figure 7. Age distribution of 8th and 9th graders, CAPS Wave 1, 2002

