Preliminary draft

Schooling as a Lottery: Racial Differences in School Advancement in Urban South Africa

David Lam
University of Michigan
davidl@umich.edu

Cally Ardington
University of Cape Town
Cally.ardington@uct.ac.za

Murray Leibbrandt
University of Cape Town
Murray.leibbrandt@uct.ac.za

Date of draft: March 1, 2007

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 (SALDRU) at the University of Cape Town. Cally Ardington is Lecturer in the Department of Statistics at the University of Cape Town and a SALDRU Research Associate. Support for this research was provided by the U.S. National Institute of Child Health and Human Development (R01HD045581), the Fogarty International Center of the U.S. National Institutes of Health (D43TW000657), and the Mellon Foundation.

Abstract:

This paper uses the Cape Area Panel Study (CAPS), a longitudinal survey of 4,800 youth in Cape Town, South Africa, to analyze the impact of individual, household, and school characteristics on progress through school. CAPS provides data on 1,500 students who were enrolled in grades 8-10 in 2002. Following these students over the next three years, we document large differences in the probability of grade advancement between white, coloured, and African youth. Probit regressions indicate that grade advancement between 2002 and 2005 is strongly associated with the respondents' scores on the literacy and numeracy evaluation administered in CAPS in 2002. The effect of these scores is much stronger for white and coloured students than for African students, a result that is consistent with our stochastic model of grade repetition. Controlling for 2002 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 8^{th}, 9^{th}, and 10^{th} graders for the next three years. We find large racial differences in the probability that students successfully advance three grades in school between 2002 and 2005 -- 82% of white students advance three grades, compared to 42% of coloured students and only 30% of African students.

As a framework for understanding the determinants of progress through school, we develop a stochastic model of grade advancement. From the perspective of a student trying to decide whether to enroll in a given grade and how much effort to invest in school, performance in school in a given year depends on systematic components such as student ability, student effort, and inputs from home and school, as well as a stochastic component that reflects randomness in the quality of teachers, the translation of effort into measurable performance, and other features of the learning environment. We show that high variance in this stochastic component can generate an equilibrium characterized by high enrollment, low effort, and high rates of grade repetition, features that are typical of predominantly black schools in South Africa. We also show that higher variance tends to reduce the impact of variables such as ability and household income on the probability of grade advancement.

After developing our theoretical model, we analyze the determinants of grade advancement and school enrollment using a using a rich set of variables from the Cape Area Panel Study. These variables include previous school outcomes, performance on a baseline literacy and numeracy evaluation, and household level variables such as income and parental schooling. Our empirical results are highly consistent with our theoretical model. While there is a strong impact of first-period test scores and household income on the probability of advancing in school, the effect of these variables is considerably weaker for African students than for coloured and white students. We show that a small set of variables, including baseline test scores, household income, and parental education, can explain most of the large racial difference in grade advancement. The results suggest that the human capital students bring with them into high school largely determines their chances of completing high school. Although African, coloured, and white students experience vastly different levels of resources in the high schools they attend, these differences are much less important in explaining high school success than the skills the students bring into high school. Put another way, while there is appropriate concern about the large quality differences in high schools, our results suggest that even if African and white students were to attend identical high schools, there would still be large racial differences in grade advancement.

I. Historical Background and Empirical Regularities

A. South African schools and the legacy of apartheid

A series of cross-national standardized tests have shown that South African learners are not internationally competitive (Van der Berg, 2005 and Crouch and Vinjevold, 2005). Indeed South Africans are shown to have performed poorly even within even within Africa. Given that the South African population is dominated by non-white groups that were disadvantaged under apartheid, an obvious explanation for this poor performance is that it evidences a lingering legacy of educational inequities from the apartheid era. There is some *prima facie* evidence to this effect in the fact that, in the midst of South Africa's poor performance on these international tests, there are pockets of international excellence that are strongly correlated with race. Similar disparities emerge from work analyzing matriculation exam results, in particular scores in mathematics and science (Van der Berg (2005), Bhorat and Oosthuisen (2006)).

Moving beyond simple descriptions of these disparities to more detailed explanation has proven to be elusive. There is mounting evidence in recent studies that the disparities are no longer simply a problem of access to education or government budget allocations for education (Fiske and Ladd (2005), Vinjevold and Crouch (2006), Van der Berg (2005)). Indeed, in these dimensions the post-apartheid government has achieved major progress and major equalization by race. At the same time this same literature is in agreement that progress on enrollments and budget equalization has not led to the equalization of educational outcomes.[1]

Part of the explanation for this lies in the fact that budget allocations provide an aggregate view of educational equalization that masks remaining inequities at the school level. For example, the equalization of these budget shares has translated into improved but not fully equalized pupil teacher ratios and other school infrastructure variables across schools (Fiske and Ladd (2004) and (Yamauchi, 2005)). Although there are greater possibilities to exercise school choice in the post-apartheid school milieu, the constraints facing learners are such that, the majority of black South Africans are still in schools that have poorer educational infrastructural inputs.[2] Clearly then, such input inequities have to be taken into account in any analysis of the performance of South African learners.

A fair amount of work within the education production function approach has attempted to glean more detail on the role of these input inequities on educational performance (Case and Deaton, 1999; Crouch and Mabogoane, 1998, 2001; Van der Berg, 2005; and Bhorat and Oosthuisen, 2006). The overriding conclusion from these studies seems to be that even after controlling for these infrastructural indicators a large part of student performance remains unexplained. This has led van der Berg (2005) and Bhorat and Oosthuizen (2006) to speculate about harder to measure aspects of student achievement such a school management and teacher

[1] Vinjevold and Crouch (2006) bluntly and usefully re-state this situation as being one in which South Africa has succeeded in giving mass access to education but has not managed to maintain or increase quality in the process. They give a longer-run perspective to this issue by arguing that the quality of South Africa's education system has been in a long-run decline because of its inability to cope with the broadening of access to education to white South Africans from the 1950s onwards.

quality. These analyses are generally conducted with school-level data and such variables are not well measured in the school-level data sets. This makes it hard to be definitive about these factors.

In any event, it is widely recognized that school infrastructure variables will always be only part of the story. Another important long-run impact of apartheid and apartheid education in particular is the fact that it leaves black parents and black communities without the resources to create a favorable home environment for learners. There has been some attempt to control for this in school-level studies by merging community level socio-economic variables into the school-level analyses. However, as these variables are only loosely connected to any actual learner and to any school, this is not completely convincing.

Given this elusiveness there would seem to be high returns to changing the perspective and to attempt to view the post-apartheid educational milieu through the eyes of the learner in the context of their household, their community and their school. Such a view can be pursued with household surveys. We analyze a new panel study of youth that provides us with much more detailed information on young people, their households, and their communities, than is available in school-based studies. Another important strength over school-based data is that we are able to follow young people over time, whether or not they remain in school. This permits us to study, for example, whether shocks to the household such as the death of a household member have an impact on whether young people stay in school.

B. Data: The Cape Area Panel Study

This paper uses the Cape Area Panel Study (CAPS), a longitudinal study of youth and their families in metropolitan Cape Town. Details about the design of CAPS are available in Lam and Seekings (2005). Wave 1 of CAPS, which was collected in 2002, included 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 sample cluster. Cape Town has three predominant population groups – coloured, African/Black, and white. The

---

[2] Seloud and Zenou (2003) provide a useful model of this constrained optimization process that shows how hard it is for previously disadvantaged South Africans to improve their schooling through school choice. Clearly some black South Africans have been able to exercise choices that improve their school environment. Van der Berg (2005) records the beginnings of an increase in within-race inequality in terms of both inputs and outputs as one would expect given increased options for all learners.

distribution of the Cape Town population in 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. 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. We will also use data about characteristics of each respondent's schools merged from the School Register of Needs.

Table 1 shows the sample size in Wave 1 and Wave 3 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.[3] The overall rate of attrition between Wave 1 and Wave 3 was 17%, with significant differences across population groups. The African attrition rate is 21%, with most of the attrition resulting from migration back to the rural Eastern Cape province that is the main sending region for Africans living in Cape Town. The coloured

[3] As in most South African household surveys, CAPS response rates were high in African and coloured areas and low in white areas. Household response rates were 89% in African areas, 83% in coloured areas, and 46% in white areas. Young adult response rates, conditional on participation of the household, were quite high, even in white areas. Given household participation, response rates for young adults were 93% in African areas, 88% in coloured areas, and 86% in white areas (Lam and Seekings, 2005).

population has its roots primarily in Cape Town, a factor contributing to its lower 10% attrition rate. The 23% attrition rate for whites includes both migration out of Cape Town (including migration out of South Africa) and a significant number of refusals.

A major focus of this paper is the comparison of schooling outcomes for African, coloured, and white youths. These three population groups were subject to very different treatment under apartheid. Many of these apartheid-era differences are likely to continue affecting 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, occupied an intermediate status under apartheid, with higher expenditures on schooling, fewer restrictions on residential mobility, and better access to jobs.

C. School enrollment, grade repetition, and work

This section provides an overview of key patterns in school enrollment, grade attainment, grade repetition, and labor force activity that form the backdrop for the school transitions we will analyze below. 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; 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 80% of Africans were in school at age 8, compared to 99% for coloured and white 8-year-olds. 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.

The second panel of Figure 1 shows the number of grades completed at each age for our 20-22 year-old Wave 1 respondents. The figure 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, 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.

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. Coloured and African students both fail grades at a much higher rate that whites, with higher failure rates for males. African and coloured males have failed an average of one grade by age 17. Taken together, the three panels in Figure 1 document a school environment characterized by almost universal primary education, high enrollment rates up to at least age 16, with grade repetition playing a large role in explaining the racial gap in schooling. Africans have particularly high rates of grade repetition, combined with high enrollment rates into the late teenage years.

While this paper focuses on schooling, it is important to keep in mind the labor market environment faced by youth during and after the school-age years. 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 and by the expected impact of schooling on wages and employment. Table 2 shows the percentage of young people who did any work for pay or family gain during the 12 months prior to the CAPS Wave 1 survey in 2002, broken down by race, age, and sex. Work is defined broadly, and includes any work done during the year. This includes work during school vacations, so it is important to keep in mind that work is not necessarily directly competing with school. As the table shows, there are enormous differences in the work experience across racial groups. At age 17, over half of white males and females report having worked in the last year, compared to only 1% of African females and 7% of African males. Coloured youth are in between, with 26% of both males and females having worked in the last year at age 17. At age 22 only 24% of African female and 35% of African males report having worked in the last year, compared to over 75% of the other four gender/race groups.

Summarizing the patterns in Figure 1 and Table 2, we see that African teenagers in Cape Town tend to have high rates of school enrollment, high rates of grade repetition, and low rates of employment. These patterns are very similar to those that would be found for African youth in all of South Africa (Anderson et al., 2002). Limited labor market opportunities, driven in part by extreme spatial segregation that is a legacy of apartheid, presumably plays an important role in explaining both the low employment and the high school enrollment. Coloured youth have significantly higher employment rates than African youth, a possible reflection of both closer geographic proximity to jobs and the legacy of the coloured labor preferences that existed in the Western Cape under apartheid. There appears to be more of a tradeoff between school enrollment and work among coloured youth, especially for males. Whites have both the highest rates of employment and the highest levels of school enrollment and schooling attainment, an indication that work and school in the teenage years are not entirely incompatible.

2. 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. We focus in particular on the combination of high enrollment rates and high rates of grade repetition documented in Figure 1. Making schooling decisions in an environment with high failure rates may lead to a number of important outcomes. One is that students and their parents may find it difficult to predict their probability of success in a given school year. Given large crowded schools with limited resources, it is hard for students and parents to control or predict key inputs such as the quality of their teachers, the ability of their classmates, or the financial status of the school. In addition to the fact that learning is compromised in such an environment, it is also likely that the evaluation of learning is imperfect. While students everywhere tend to rationalize failure as the result of bad luck, there may be more truth to these perceptions in poor schools with high failure rates.

To model this environment, consider a simple stochastic model of grade advancement. Suppose that students are evaluated at the end of the school year based on a final score *S*. This score is an imperfect summary of the students actual knowledge at the end of the year, which can be characterized by a learning production function $K=F(\mathbf{X})$, where **X** is a vector that includes all of the inputs that affect knowledge at the end of the year, including the student's previous knowledge, the effort put into school, school inputs, and family background characteristics such as parental schooling. The score for student *i* also includes a stochastic component u_i reflecting a wide variety of reasons for discrepancies between knowledge and scores. In the most literal sense, these could include errors in the marking of exams and recording of grades. More broadly they include any problem in the educational environment that causes real learning to be unrewarded in the evaluation of students to determine advancement. For example, weak teachers in bad schools may teach and test in such a poorly organized way that mastery of certain course material has no impact on final evaluations. Assume that we can summarize this with a linear model

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

where **X** is a vector representing the systematic determinants of student performance and *u* is a stochastic component that is uncorrelated with the variables in **X**. We assume that there are a large number of independent components in *u*, making it reasonable to assume that it is normally distributed, $u \sim N(0,\sigma)$. We will assume that the perceptions of students and their parents about the process driving Equation (1) are consistent with reality, but for purposes of understanding decisions about enrollment and effort it is only necessary to assume that Equation (1) describes those perceptions, whether or not they are correct.

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

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

where Φ is the cumulative of the standard normal distribution.

A. 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 **X**, such as previously acquired human capital, parental education, or a measure of school quality. Denote this variable by X_1, and its corresponding coefficient in Equation (1) by β_1 (we will assume that $\beta_1 > 0$ 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_1 on the probability of passing, we differentiate Equation (2) to get:

$$\frac{\partial P(S_i > T)}{\partial X_{1i}} = \frac{\beta_1}{\sigma}\phi\left(\frac{T - \boldsymbol{\beta}'\mathbf{X}_i}{\sigma}\right) = \beta_1 f\left(T - \boldsymbol{\beta}'\mathbf{X}_i\right), \tag{3}$$

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_1 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_1 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 $g_i = T - \boldsymbol{\beta}'\mathbf{X}_i$ as the gap between the deterministic component of the score and the threshold, and taking the derivative of Equation (3) with respect to σ,

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

The cross-partial derivative in Equation (4) is negative when $g_i < \sigma$ and positive when $g_i > \sigma$.

Figure 2 illustrates several features of the model. The top panel shows the density for four normal distributions with mean zero, with standard deviations equal to 1, 1.1, 2, and 2.1. The bottom panel shows the cumulative distribution function for the same distributions. Point A in the top panel, which sits at one standard deviation for the distribution with σ=1, is one of two fixed points of the density when σ is increased (the other being -1σ). An increase in the standard deviation causes a decrease in the density in the range within one standard deviation of the mean and an increase in the density outside of that range. Point B shows that the density rotates around 2 when σ=2. This implies that 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. These could be represented by the distributions with σ=1 and σ=2 in Figure 2. Suppose both students are exactly at the passing threshold based on the deterministic component, $T - \boldsymbol{\beta}'\mathbf{X}_2 = T - \boldsymbol{\beta}'\mathbf{X}_1 = 0$, implying that they will pass if the draw is positive and fail if the draw is negative. Then looking at the densities in Figure 2 at 0, and looking at Equation (4), we see that the marginal effect of ability on the probability of passing will be twice as high for the student in the low-variance environment. Put another way, a one point increase in the deterministic component of the grade will have twice as large an impact on the probability of passing 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.[4] 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 of characteristics on the probability of passing. For a student at the bottom of the distribution, a one point increase in her grade would have almost no chance of pushing her over the passing threshold in a low-variance regime. But in a high-variance regime there is a greater chance that her one additional point will be matched by a high positive draw from the stochastic component, pushing her over the threshold.

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 some characteristic, the regression gives 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 the characteristic on human capital accumulation (differences in β) and differences in the variance in the human capital production function (differences in σ).

B. How hard do students work?

One set of variables in the learning 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

[4] 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.

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 (these should 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. We will thus expect less effort from students who are within one standard deviation of the passing threshold if we increase the stochastic variance.

C. Who goes to school?

Assuming that school enrollment is 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. Taking the derivative of Equation (1) with respect to σ,

$$\frac{\partial P(S_i > T)}{\partial \sigma} = \left(\frac{T - \boldsymbol{\beta}'\mathbf{X}_i}{\sigma^2}\right)\phi\left(\frac{T - \boldsymbol{\beta}'\mathbf{X}_i}{\sigma}\right), \qquad (5)$$

The probability of passing increases with σ for those who would have failed based on the deterministic component, and decreases for those who would otherwise have passed.

As σ increases, the expected probability of passing the grade is increasingly determined by the stochastic component. Consider two groups of students, a low-skilled group for whom $T > \boldsymbol{\beta}'\mathbf{X}_1$ and a high-skilled group for whom $T < \boldsymbol{\beta}'\mathbf{X}_2$. 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*.

The point is illustrated in Figure 2. The CDF in the bottom panel shows the probability of passing for different values of $\boldsymbol{\beta}'\mathbf{X}_i - T$, with students exactly at the passing threshold having 50% probability of passing. Suppose that given the opportunity cost of being in school, students only enroll if they have a 30% chance of passing. If σ=1, only those to the right of the line at Point C will choose to enroll. Following the line to the top panel, we see the marginal impact of characteristics for the lowest scoring students. If we increase the standard deviation to σ=2, the line dividing those with 30% probability of passing shifts to the left to Point D. Looking at the distribution of the two densities to the right of Points C and D, it is clear that the average return to characteristics (as given by the height of the density) will tend to be lower in the distribution with σ=2, although a precise statement about this depends on the distribution of the deterministic component.

D. Effects over multiple years

Suppose we follow students over more than one year of school. We can generalize (1) by adding a subscript t and making assumptions about the correlation of the stochastic term across years. The simplest case is to assume that u_{t+k} is uncorrelated with u_t for all k, an assumption that fits with our characterization of the idiosyncratic and unpredictable nature of the stochastic component. The probability of passing all years from year 1 to year *n* is

$P_{i,1n} = \prod_{t=1}^{n} P_{i,t} = \prod_{t=1}^{n} P(S_{i,t} > T_t)$. Consider the simple case in which **X**, **β**, σ, and T are the same in every year, implying that students are moving through a series of grades in which a given student would get the same score in every year in the absence of the stochastic component. If the stochastic terms are uncorrelated across years, then the probability of passing is identical in every year, $P_1 = P_t = P_n$. In looking at the impact of characteristics on the probability of passing all *n* grades, it is helpful to take logs and consider the proportional impact, since the absolute impact will depend on the absolute level of the probability, which will decrease as we expand the number of years. Taking the derivative of $\ln(P_{i,1n})$ with respect to some characteristic X_1, and recalling the result from (3), we get

$$\frac{\partial \ln P_{i,1n}}{\partial X_{1i}} = \frac{n\beta_1}{\sigma P_t}\phi\left(\frac{T - \boldsymbol{\beta}'\mathbf{X}_i}{\sigma}\right), \qquad (6)$$

where P_t is the probability of passing in any single year. Note that this is equivalent to (3) in the case where n=1 and we take the derivative of the log. The derivative in (6) shows that the proportional impact of characteristics on the probability of passing increases as we track progress across more grades. If one additional IQ point gives a student a 2% higher probability of passing a single grade, then that IQ point will give the student a 10% higher probability of passing five consecutive grades.[5] While (6) is derived for the simple case in which every grade is identical, the result is quite general as long as we assume that the stochastic component is uncorrelated across years. The stochastic component introduces noise into each year's results, causing some poor students to pass over better students and weakening the link between ability (for example) and scores. Over multiple years the better students emerge with a clearer advantage, however, as the systematic component dominates the uncorrelated stochastic component. More generally, we could imagine that there are some components of the stochastic term that are correlated over time, and other components that are uncorrelated. The uncorrelated components, which in our model represent the pure noise in the link between learning and measurement of performance, become less important when we look across a larger number of years. This implies that the

[5] More accurately, a .02 difference in log probabilities in one year will imply a .10 difference in log probabilities in five years.

proportional impact of characteristics on pass rates will be larger when we look at passing over multiple years.

E. Impact of failing on enrollment and future success

Another interesting implication of the model is that the impact of failing grades on future enrollment will in general depend on the magnitude of the stochastic component. If we assume that students have some uncertainty about their own ability and likelihood of future success in school, then scores received each year (and decisions about promotion) are important signals about that ability. A larger stochastic component in scoring implies that grade promotion is a noisier signal about the student's ability and future probability of success. We may expect, then, that past failure will be a weaker predictor of future enrollment in a regime with high perceived variance. Past failures will also be a weaker predictor of future probabilities of passing in the high-variance environment, since high variance weakens the link between failure and actual learning. While it is perceived variance that affects the link between past failure and future enrollment decisions, the actual variance will be important for the link between past failure and future promotion. The perceived variance may also play an additional role by affecting the impact of past failure on future effort, which in turn will affect future probabilities of passing.

F. Externally evaluated standardized exams

An important feature of the South African school system is that there is a nationally standardized matriculation exam in Grade 12, with external evaluation of exams. Performance on this exam has important consequences for both students and schools, with extensive national media coverage of matriculation pass rates when they are announced in December of each year. Preparation for the matric exam is a major focus of student effort during Grade 12. The matric exam provides an interesting test of our model, since it implies that passing Grade 12 will have somewhat different features than passing Grades 8-11. If the standardization and external evaluation lead to both a reduction in variance and smaller differences in sigma across different population groups, then we would expect there to be a larger impact of characteristics on pass rates, and smaller differences across population groups in the impact of characteristics.

While the cases discussed above are highly stylized, some important realism is captured by the model. The most important is that it is relatively easy to generate an equilibrium 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 predominantly black schools in the South African educational system. 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. Most importantly, both the probability of grade advancement and the probability of enrollment will tend to be less affected by characteristics such as ability and family background in an environment in which there is a larger stochastic component to measured performance. Second, the impact of failing grades on future enrollment and grade advancement will be lower when there is a high stochastic component. Third, the impact of characteristics on passing will be larger when we look at passing over multiple grades than when we look at passing a single grade. Fourth, the impact of characteristics will be larger and the differences between population groups will be smaller for passing the standardized Grade 12 matriculation exam than for passing earlier grades.

3. Empirical Evidence

In this section we analyze the extent to which empirical evidence on progress through school is consistent with our stochastic model of grade repetition. We begin with a descriptive overview of grade progression for our sample of 8^{th}, 9^{th}, and 10^{th} graders in 2002. We then estimate regressions to test some of the specific predictions of the model.

A. 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 8^{th} and 9^{th} grade students would have reached grades 11 and 12, respectively, by the 2005 wave. Those in 10^{th} grade in 2002 who stayed in school and passed the 10^{th} and 11^{th} 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.

Table 3 shows the activities in 2005 of those where 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.

B. 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 CAPS 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 respondents 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 took the test in a second language, while white and coloured students took 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 3 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 3 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 4 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 4 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 5 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.

C. Probit regression results

1. Regressions for 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. One of our key empirical questions is whether there are differences in the impact of individual and household characteristics on grade advancement and school enrollment. In addition to the possibility that the coefficients that map characteristics into school performance may differ across racial groups, we have argued above that the stochastic component in school outcomes may also differ across racial groups. We hypothesize that we will either have different coefficients across population groups, or different variances in the stochastic component, or potentially both. Different variances imply heteroscedascity of a particular type, although in practice we cannot expect to distinguish between differences in coefficients and differences in the variance when estimating probits. As in Cameron and Heckman (2001), who estimate separate models for whites, blacks, and Hispanics in U.S. data, we assume from the outset that we should estimate separate regressions for each of our three population groups – African, coloured, and white. For each coefficient and for each pairwise combination of races we test for equality of the coefficients.

Our first set of probit regressions look at progress through school. Means of the dependent and independent variables are presented in Table 4. The dependent variable is equal to 1 if the respondent advanced a full three years in school by 2005 – 8th graders reached at least grade 11, 9th graders reached at least grade 12, and 10th graders successfully completed 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. As shown in Table 4, the percentage of students advancing three grades varies enormously by race, 30% for Africans, 41% for coloureds, and 82% for whites. Independent variables include a dummy for female and dummies for being in 8th and 9th grade in 2002 (10th grade omitted). We use the number of grades failed by 2002 and the standardized LNE score as measures of school performance and learning at the time of the Wave 1 survey. Household characteristics include the log of per capita household income, an indicator of household shocks between 2002 and 2005, years of schooling of the mother and father, and dummy variables for missing parental schooling. We also include the age-sex-specific unemployment for individuals with less than 12 years of schooling in the census sub-place and the student-teacher ratio in the student's 2002 school.

Table 5 presents the first set of probit regressions. We estimate very large effects of the LNE score and the number of previous grades failed, demonstrating the importance of previous learning and performance in school. The magnitude of these effects will be discussed below. These effects differ substantially across races. Previous grades failed has a much less negative effect on grade advancement for African students than for coloured and white students, and the LNE score has a smaller positive effect for African students. The impact of household income is not statistically significantly for Africans, but is strongly positive for coloured and white students. As shown in Column 4, we strongly reject equality of the African and coloured coefficients for number of grades failed, LNE scores, and household income. The small white sample leads to large standard errors on the white coefficients, making it impossible to reject equality of the African and white coefficients on LNE scores and income, in spite of large differences in the point estimates. We do reject equality of the African and white coefficients on number of grades failed.

The predicted impact of the LNE scores and household income can be seen in the predicted probabilities graphed in Figure 6. The top left panel shows the predicted probability of advancing three grades as a function of LNE scores, based on the coefficients in the three separate probits by race. The predicted values are calculated for a female in grade 8 in 2002 with mean income, mean unemployment rate, and parents' education equal to 10 years. There is a very strong effect of the LNE scores on grade advancement, with the steepest slope around an

LNE score of zero for coloured students (in addition to zero being the overall mean for the sample, it is very close to the mean for the coloured subsample). The slope for Africans is flatter over most of the range between -2 and +2 standard deviations. Referring back to our theoretical model, this flatter slope could result from either a smaller coefficient on LNE scores in the underlying learning production function or from a higher variance in the stochastic component of measured performance (or some combination of the two).

The upper right panel of Figure 6 shows a similar pattern for the impact of income on grade advancement. The impact of income is much smaller for Africans than for white or coloured students. This result may seem surprising, since we might expect to find large effects of income over the range of income covered by the African sub-sample. The poorest part of the African sample is in deep poverty, while the upper tail of the distribution has levels of income that should make it much easier to keep children in school and provide them with basic inputs to support their progress through school. Our interpretation of this low impact of income on African grade advancement is that it is a symptom of the inefficient and chaotic school environment, which is ineffective in translating either higher ability or better resources into measurable improvements in school performance.

Figure 6 demonstrates another important point that will be explored in greater detail below. The predicted probability of advancing three grades is higher for Africans than for coloureds over a considerable range of test scores and income at the bottom of the distribution. This includes comparisons at zero, the mean of both the LNE score and the income variable. As we will see below in our counterfactuals, our results imply that the gap between African and coloured students in grade advancement is more than fully explained by differences in characteristics, with test scores and income explaining most of the difference. Even the enormous gap between Africans and whites can be mostly explained by differences in the variables included in the regressions in Table 3.

Looking at other variables in our probit in Table 5, parental schooling has surprisingly weak effects on grade advancement, with none of the coefficients statistically significant at conventional levels. For Africans these coefficients continue to be insignificant even when the LNE scores, number of grades failed, and student-teacher ratio are omitted. For coloured students we estimate a significant positive effect of father's schooling when the previous performance outcomes are omitted. Negative household shocks such as the death or job loss of a household member have significant negative effects on grade advancement for both Africans and whites, though not for coloureds.

The neighborhood unemployment rate is not significant for any racial group. It is included here as an attempt to capture two possible effects. On the one hand, the opportunity cost of time may affect either effort in school or the probability of dropping out. On the other hand, better employment prospects might stimulate young people to stay in school and work harder in school. These effects may be cancelling out in our data, although it is also possible that our use of census subplaces does not capture the appropriate labor market. While white and coloured youth appear to have much better job opportunities than African youth due to geographical proximity, family networks, and language skills, there may not be sufficient geographical variation in job opportunities within racial groups to identify an effect. The learner/educator ratio in the students 2002 school, incorporated into our data from the South Africa School Register of Needs, has a negative impact on grade advancement for African students, but not for coloured or white students.

The specification in Table 5 includes a number of variables that are endogenous with respect to either previous school performance or contemporaneous school choice. For example, we may be concerned that the impact of parental schooling is low in Table 5 because we have controlled for the student's test scores and number of previous grades failed. Most of the effect of parental schooling may work through those earlier outcomes. Similarly, controlling for school quality may weaken the impact of income or parental schooling, since much of the effect may work through the choice of schools. We have estimated many alternative specifications to explore these issues.

2. Regressions for school enrollment

Table 6 presents regressions in which the dependent variable is school enrollment in 2004. We use 2004 enrollment because all our Wave 1 8th, 9th, and 10th graders should have been in school in 2004 in order to make normal progress through secondary school. We include a dummy variable for whether they failed their grade in 2002 in order to see whether students drop out or return to school after failing a grade. The other variables in Table 6 are the same as those in Table 5. We exclude whites from these regressions because over 95% of whites are enrolled in 2004, making it difficult to estimate meaningful regressions given the small white sample.

As in Table 5, we estimate significant negative effects of the number of previous grades failed and significant positive effects of the LNE score. As in Table 5, the point estimates for the effects of these variables are larger in magnitude for coloured students than for African students, though in neither case can we reject equality of the coefficients. The estimated effect of household income on enrollment is statistically insignificant for Africans, but is strongly positive for coloureds. While many will find it surprising that income does not affect enrollment for Africans, we interpret it as indicating that the combination of very low opportunity cost, high returns to schooling, and imperfect measuring of performance make the benefits of being enrolled sufficient to overcome the direct out-of-pocket expenses such as fees and uniforms. Failing the grade in 2002 has a strong negative effect on enrollment in 2004 for both Africans and coloureds, but the effect is much greater for coloured students. This is consistent with our interpretation of the responses of Africans and coloureds to differences in the school environment. Failing a grade prior to 2002 is a weaker predictor of future success in school for Africans than for coloureds. Consistent with this fact, Africans are less likely to drop out if they fail their grade in 2002.

The bottom two panels of Figure 6 plot the predicted enrollment as a function of LNE scores and per capita household income for Africans and coloureds, with separate predictions for those who failed and those who did not fail in 2002. Several features of the graphs are worth noting. First, the lines for Africans are much flatter than the lines for coloureds, showing the much weaker responsiveness of enrollment to ability or income for Africans. Second, the predicted enrollment is higher for Africans over a very broad range of LNE scores and income. Even Africans who failed their grade in 2002 have a higher predicted probability of being enrolled in 2004 than coloured students who passed in 2002 over much of the low range of LNE scores. Finally, we see that the gap in predicted enrollment between those who passed in 2002 and those who failed in 2002 is much larger for coloured students. Taken together, we see an equilibrium for Africans that is characterized by high enrollments that are only weakly related to ability, previous performance, or household income.

D. Explaining gaps in grade advancement and enrollment

Table 7 uses the probits from Tables 5 and 6 for counterfactuals designed to estimate the extent to which racial differences in characteristics explain differences in grade advancement and enrollment. The approach is very similar to the approach of Cameron and Heckman (2001). Because the coefficients are often very different between groups, we do the counterfactuals using each racial group as the baseline for any given pairwise comparison. Looking at column 1, we see that the actual gap between African and coloured students in the probability of advancing three grades between 2002 and 2005 is 11 percentage points. The gap in the predicted values from the African and coloured regressions is 11.4 percentage points, almost identical to the actual gap. The first counterfactual assumes that Africans have the same covariates as coloureds (that is, the African coefficients are applied to coloureds). This counterfactual predicts a gap of -0.057, implying that Africans would have had 5.7 percentage point *higher* probability of advancing three grades if they had the characteristics of coloured students. In other words, we more than fully explain the gap between African and coloured students when we equalize their characteristics. When we combine African characteristics with coloured coefficients we predict an even larger gap in favor of Africans, with coloured students having a 12.5 percentage point lower predicted probability of grade advancement.

Looking at the African-white comparisons in column 2 of Table 7, the actual gap in the probability of advancing three grades between 2002 and 2005 is 50.7 percentage points. Taking the predicted values from our probit regressions, we predict a gap of 48.4 percentage points. In the first counterfactual, which assumes that Africans have the same covariates as whites, the predicted gap between Africans and whites drops to -5.5 percentage points. As was the case with the African-coloured comparison, we more than fully explain the difference in grade advancement between African and white students when we equalize their characteristics. Doing the counterfactual in the other direction, assuming that whites have the same covariates as blacks, we get a very similar result, with a predicted gap of -9.4 percentage points. The last two rows use regressions that exclude the LNE scores. These also explain most of the gap in grade advancement.

Table 8 calculates predicted values using individual coefficients. Following Cameron and

Heckman (2001), we assign mean values from the other racial group for a given characteristic. For example, to see the impact of giving Africans the LNE scores of coloured students, we give all African students the mean coloured LNE score, using the African regression and keeping all other characteristics of Africans unchanged. Looking at column 1, we see that assigning coloured LNE scores to Africans would reduce the coloured-African gap in grade advancement by 6.5 percentage points (out of a total gap of 11.0). Column 2 shows that if we use the coloured regressions and give coloured students the LNE scores of Africans, their probability of advancement would fall by 13.3 percentage points, more than the 11.0 percentage point coloured advantage. Column 5 shows that giving Africans the LNE scores of white students would lower the African-white gap by 19.6 percentage points (out of 50.7).

Equalizing log per capita household income also has a large impact on the racial gap in grade advancement. The effect is larger when the coloured and white coefficients are used, since those coefficients are much larger than the African coefficients. Column 2 shows that giving the African mean income to coloured students, using the coloured regressions, lowers their probability of advancement by 10 percentage points, 92% of the African-coloured gap. Giving African mean income to white students, using the white regressions, lowers their probability of grade advancement by 28 percentage points, 55% of the African-white gap. The last row of Table 8 shows the impact of equalizing both the LNE scores and the number of grades behind in 2002, leaving all other characteristics unchanged. These counterfactuals show the importance of initial schooling achievement in predicting progress through secondary school. These two variables alone explain from 80% to 150% of the African-coloured gap and from 50% to 77% of the African-white gap in grade advancement. These variables are themselves an indicator of a large number of factors that will have affected previous schooling outcomes, including school quality, household characteristics, and student's ability. While we cannot be sure exactly what caused the large racial gaps in initial test scores and grade attainment, the important point is that students entered secondary school with large pre-existing achievement gaps. Our results suggest that it would be very difficult to equalize the probability of advancing through secondary school without reducing these initial differences.

4. Conclusions

Following young people over a three-year period in the Cape Area Panel Study, we document large racial differences in the probability that 8th, 9th, and 10th graders make normal progress through school. While 82% of white students in the 8th, 9th, and 10th grade in 2002 had reached grade 11, grade 12, or completed grade 12, respectively, by 2005, only 30% of African students and 42% of coloured students had made the same progress. While dropping out and grade repetition both contribute to the gap, grade repetition is the most important factor, especially for Africans. A simple description of the school environment for African secondary school students is that it is characterized by high enrollment rates and high rates of grade repetition, with many students falling at least one grade behind in school.

Our theoretical model demonstrates that a large stochastic component to grade advancement can have important effects on who attends school, how much effort they invest in school, and how individual and household characteristics affect the probability of grade advancement. We show that by increasing the variance in the stochastic component of grade advancement we can generate an equilibrium that looks very much like African schools in Cape Town – high rates of enrollment, low levels of effort, and high rates of grade repetition. This model also implies that characteristics such as parental income and previous school performance will have a lower impact in African schools than in coloured or white schools, assuming that African schools have a larger stochastic component in grade advancement.

The results of our probit regressions are highly consistent with our theoretical model. While we find a strong effect of test scores and household income on the probability of grade advancement for all races, the effect of these variables is significantly weaker for African students. While this could indicate that there is an interaction between school quality and other inputs, it is also consistent with a higher variance in the random components of grade advancement in African schools. This high variance helps explain the high school enrollment among African students, even in the face of high failure rates. For these students, high school has elements of a lottery, with even low-ability students having an incentive to be enrolled.

Our estimates show that household income and indicators of previous achievement such as test scores and the number of grades behind in 2002 are strong predictors of subsequent progress through school. Estimating counterfactuals using our separate probit regressions for each race, we find that we can entirely explain the racial gaps in grade achievement between 2002 and 2005 by differences in the characteristics of students and their households in 2002. Taken at face value, the results suggest that eliminating the large racial differences in the quality of secondary schools would have very limited impact on the racial gap in grade progression. African students begin secondary school with such large disadvantages in terms of test scores and previous school performance that it is very difficult for them to complete school at the same rate as coloured or white students. From a policy perspective, one interpretation of the results is that bad secondary schools in African areas are not the primary cause of the poor performance of African students. This interpretation also leads to an important policy challenge, however, since it implies that the problems begin much earlier than secondary school, potentially shifting the focus to poor quality primary schools and the disadvantages of growing up in poor households.

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.

Bhorat, Haroon and Oosthuizen, Morne (2006) "Determinants of Grade 12 Pass Rates in the post-*Apartheid* South African Schooling System, DPRU Working Paper, University of Cape Town.

Cameron, Stephen V., and James J. Heckman (2001) "The Dynamics of Educational Attainment for Black, Hispanic, and White Males," *Journal of Political Economy*, 109(3): 455-499.

Case, Anne, and Angus Deaton (1999) "School Inputs and Education Outcomes in South Africa," *Quarterly Journal of Economics*, August, 114(3): 1047-84.

Crouch, Luis and T.Mabogoane (2001) "No magic bullets, just tracer bullets: The role of learning resources, social advantage and education management in improving the performance of South African schools," *Social Dynamics*l 27(1):60-78.

Crouch, Luis and T.Mabogoane (1998) "When the Residuals Matter More than the Coefficients: An Educational Perspective," *Studies in Economics and Econometrics* 22(2)

Crouch, Luis and Vinjevold, Penny (2006) South Africa: Access before quality, and what to do now? Unpublished Paper, RTI, North Carolina.

Fiske, Edward and Helen Ladd, H (2004) *Elusive equity: Education reform in post-apartheid*, Brookings Institution, Washington, D.C.

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.

Selod, H and Zenou, Y (2003) Private versus public schools in post-apartheid South African cities: Theory and policy implications. *Journal of Development Economics*. 71(2): 351-394

Thomas, Duncan (1996) "Education across Generations in South Africa," *American Economic Review*, 86(2): 330-334.

van der Berg, Servaas (2005) "Apartheid's enduring legacy: Inequalities in education," Paper delivered to the Oxford University/University of Stellenbosch conference on *The South African Economic Policy under Democracy: A 10 year review*, Stellenbosch, October 2005.

Yamauchi, Futoshi (2005). Race, equity and public schools in post-apartheid South Africa: Equal opportunity for all kids. *Economics of Education Review* (24): 213-33.

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

	CAPS Wave 1, 2002			CAPS Wave 3 - 2005			
Population Group	Sample Size	Unweighted Percent	Weighted Percent	Sample Size	Unweighted Percent	Weighted Percent	Rate of Attrition
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. Percentage who worked in last 12 months, CAPS respondents in Wave 1, 2002

	African		Coloured		White	
Age	Female	Male	Female	Male	Female	Male
14	0.0	0.7	7.4	19.7	9.0	30.3
15	0.0	0.8	12.7	10.5	27.1	33.3
16	1.6	5.3	14.9	27.2	44.8	32.0
17	1.3	6.6	26.4	26.6	53.9	51.0
18	1.9	9.5	32.0	47.0	53.3	73.6
19	6.9	10.8	52.3	62.7	70.2	72.6
20	16.7	24.7	63.9	83.5	82.9	80.5
21	19.8	26.9	65.1	82.4	78.8	89.1
22	23.9	35.3	77.4	78.1	75.7	87.9
Sample Size	1,219	927	1,077	925	313	284

Table 3. Percentage in each grade or non-enrollment status in 2005, CAPS respondents in grades 8 and 9 in 2002

	8th grade in 2002			9th grade in 2002		
Status in 2005	African	Coloured	White	African	Coloured	White
Enrolled in grade 8	0.7	0.0	0.0	--	--	--
Enrolled in grade 9	6.3	2.2	0.0	0.5	0.0	0.0
Enrolled in grade 10	37.4	13.2	0.0	10.6	3.5	0.0
Enrolled in grade 11	33.4	45.7	85.3	29.5	15.5	8.0
Enrolled in grade 12	2.5	0.7	7.3	27.5	40.7	80.8
Post-secondary	1.6	1.9	0.0	1.3	1.8	3.7
Total enrolled	81.9	63.6	92.5	69.3	61.5	92.4
Not enrolled/not working	15.0	24.0	3.0	22.4	22.8	1.1
Not enrolled/working	3.1	12.5	4.5	8.3	15.7	6.5
Sample size	141	132	41	248	228	58

Table 4. Descriptive statistics, CAPS respondents enrolled in Grades 8, 9 and 10 in 2002 and observed again in 2005

	African			Coloured			White		
Variable	N	Mean	Std. Dev.	N	Mean	Std. Dev.	N	Mean	Std. Dev.
Enrolled in 2004	579	0.850	0.358	540	0.748	0.434	144	0.951	0.216
Advance 3 grades by 2005	584	0.301	0.459	545	0.415	0.493	144	0.819	0.386
Enrolled in grade 8 in 2002	584	0.241	0.428	545	0.242	0.429	144	0.285	0.453
Enrolled in grade 9 in 2002	584	0.425	0.495	545	0.418	0.494	144	0.403	0.492
Female	584	0.557	0.497	545	0.536	0.499	144	0.500	0.502
Grades failed by 2002	584	0.745	0.903	545	0.519	0.750	144	0.208	0.500
Standardized LNE total score	580	-0.544	0.786	542	0.056	0.696	143	1.180	0.550
Log per cap hh income (mean zero)	568	-0.668	0.881	509	0.273	0.857	124	1.937	0.771
Mother's education	524	8.376	2.841	492	8.697	2.643	143	12.517	1.819
Father's education	339	7.372	3.628	371	8.841	3.082	133	13.045	2.117
Mother's education missing	584	0.103	0.304	545	0.097	0.297	144	0.007	0.083
Father's education missing	584	0.420	0.494	545	0.319	0.467	144	0.076	0.267
Household shock 2002-2005	584	0.226	0.419	545	0.150	0.358	144	0.056	0.230
Local unemployment rate for age & sex	584	0.805	0.128	536	0.625	0.188	113	0.280	0.268
Learner educator ratio	540	32.424	3.779	536	30.172	3.710	133	22.946	4.069
Failed grade in 2002	569	0.178	0.382	539	0.173	0.378	143	0.035	0.184
former Dept. of Education and Training school (African)	540	0.778	0.416	536	0.024	0.154	130	0.008	0.088
former House of Assembly school (White)	540	0.024	0.153	536	0.086	0.280	130	0.938	0.241
former House of Representatives school (Coloured)	540	0.109	0.312	536	0.860	0.347	130	0.008	0.088
New School since 1994	540	0.089	0.285	536	0.007	0.086	130	0.046	0.211

Table 5. Probit regressions for probability of advancing 3 grades between 2002 and 2005, CAPS respondents in grades 8, 9 or 10 in 2002						
	Probits for grade advancement			Tests for equality of coefficients		
Variable	African	Coloured	White	African-Coloured	African-White	Coloured-White
	(1)	(2)	(3)	(4)	(5)	(6)
Grade 8 in 2002	0.355	0.484	1.2	0.261	2.669	1.825
	[0.157]**	[0.199]**	[0.492]**	(0.609)	(0.103)	(0.177)
Grade 9 in 2002	0.024	0.142	-0.532	0.350	1.436	2.108
	[0.140]	[0.148]	[0.441]	(0.554)	(0.231)	(0.147)
Female	-0.119	0.268	0.041	4.980	0.078	0.161
	[0.133]	[0.122]**	[0.552]	(0.026)**	(0.780)	(0.688)
Number of grades failed, Wave 1	-0.287	-0.683	-1.323	6.377	8.752	3.055
	[0.088]***	[0.130]***	[0.340]***	(0.012)**	(0.003)***	(0.081)*
Standardized LNE total score	0.329	0.84	0.559	9.688	0.343	0.477
	[0.082]***	[0.143]***	[0.382]	(0.002)***	(0.558)	(0.490)
Log hh income per cap.	0.12	0.415	0.507	5.613	2.576	0.137
	[0.075]	[0.097]***	[0.229]**	(0.018)**	(0.109)	(0.710)
Mother's schooling	0.018	0.038	-0.105	0.202	0.675	0.904
	[0.032]	[0.029]	[0.147]	(0.652)	(0.411)	(0.342)
Mother's schooling missing	0.235	-0.052		0.330		
	[0.376]	[0.332]		(0.565)		
Father's schooling	0.009	0.039	-0.06	0.749	0.375	0.765
	[0.021]	[0.029]	[0.109]	(0.387)	(0.540)	(0.382)
Father's schooling missing	-0.016	0.395	-1.436	1.414		
	[0.194]	[0.280]	[1.563]	(0.235)		
Household shock between 2002 and 2005	-0.265	-0.196	-1.591	0.085	4.229	4.582
	[0.146]*	[0.185]	[0.625]**	(0.769)	(0.040)**	(0.033)**
Local unemployment rate for age and sex	0.092	0.046	-0.152	0.004	0.055	0.044
	[0.645]	[0.346]	[0.877]	(0.949)	(0.813)	(0.833)
Learner/educator ratio	-0.052	0.012	-0.004	6.236	0.696	0.068
	[0.016]***	[0.020]	[0.056]	(0.013)**	(0.404)	(0.793)
Constant	1.346	-1.448	2.13	6.286	0.091	1.851
	[0.795]*	[0.782]*	[2.508]	(0.012)**	(0.762)	(0.174)
Observations	522	490	90			

Robust standard errors in brackets in columns 1-3; p-value of F tests in parentheses in columns 4-6; * significant at 10%; ** significant at 5%; *** significant at 1%

Table 6. Probit regressions for probability of enrollment in 2004, CAPS respondents in grades 8, 9 or 10 in 2002

	Probits for enrollment		Test of equality of coefficients
Variable	African	Coloured	African-Coloured
	(1)	(2)	(3)
Grade 8 in 2002	0.352	0.222	0.230
	[0.187]*	[0.199]	(0.631)
Grade 9 in 2002	0.18	-0.026	0.792
	[0.163]	[0.164]	(0.374)
Female	0.364	0.197	0.602
	[0.159]**	[0.147]	(0.438)
Number of grades failed, Wave 1	-0.347	-0.411	0.273
	[0.067]***	[0.102]***	(0.601)
Standardized LNE total score	0.214	0.32	0.550
	[0.097]**	[0.105]***	(0.458)
Log hh income per cap.	-0.011	0.338	6.211
	[0.101]	[0.097]***	(0.013)**
Mother's schooling	0.056	-0.025	3.803
	[0.026]**	[0.033]	(0.052)*
Mother's schooling missing	0.594	-0.766	4.471
	[0.298]**	[0.392]*	(0.035)**
Father's schooling	0.014	0.119	7.764
	[0.030]	[0.040]***	(0.005)***
Father's schooling missing	-0.103	0.781	3.643
	[0.272]	[0.374]**	(0.057)*
Household shock between 2002 and 2005	-0.074	-0.107	0.014
	[0.177]	[0.216]	(0.904)
Local unemployment rate for age and sex	0.683	0.087	0.771
	[0.528]	[0.421]	(0.380)
Learner/educator ratio	0.004	0.027	0.761
	[0.021]	[0.016]*	(0.383)
Failed in 2002	-0.36	-1.26	14.32
	[0.172]**	[0.164]***	(0.000)***
Constant	0.177	-0.433	0.297
	[0.871]	[0.696]	(0.585)
Observations	588	516	

Robust standard errors in brackets in columns 1-3; p-value of F tests in parentheses in columns 4-6; * significant at 10%; ** significant at 5%; *** significant at 1%

Table 7. Percentage of racial gap explained by differences in covariates

	Advance 3 grades				Enrolled in 2004
X	African	African	Coloured		African
Y	Coloured	White	White		Coloured
	(1)	(2)	(3)		(4)
Actual gap	0.110	0.507	0.397		-0.110
Including LNE scores					
Predicted gap	0.114	0.484	0.370		-0.108
Gap when X has Y covariates	-0.057	-0.055	-0.117		-0.127
Gap when Y has X covariates	-0.125	-0.094	0.017		-0.264
Excluding LNE scores					
Predicted gap	0.114	0.482	0.369		-0.106
Gap when X has Y covariates	-0.016	0.009	-0.063		-0.108
Gap when Y has X covariates	-0.069	0.008	0.087		-0.224
Based on probit regressions in Tables 3 and 4					

Table 8: Predicted change in racial gap when individual covariates are changed

Variable	African equated to coloured	Coloured equated to African	Coloured equated to white	White equated to coloured	African equated to white	White equated to African
	(1)	(2)	(3)	(4)	(5)	(6)
Outcome: Advance 3 grades						
Actual Gap	**0.110**	**-0.110**	**0.397**	**-0.397**	**0.507**	**-0.507**
Change in gap due to equalizing:						
Number of grades failed, Wave 1	0.021	-0.043	0.059	-0.078	0.051	-0.145
Standardized LNE total score	0.065	-0.133	0.256	-0.125	0.196	-0.204
Log hh income per cap.	0.036	-0.101	0.183	-0.166	0.101	-0.281
Mother's schooling	0.001	-0.003	0.047	0.070	0.028	0.074
Father's schooling	0.005	-0.018	0.067	0.057	0.021	0.071
Household shock between 2002 and 2005	0.006	-0.004	0.004	-0.023	0.013	-0.044
Local unemployment rate for age and sex	-0.005	0.002	-0.004	-0.009	-0.015	-0.014
Learner/educator ratio	0.036	0.007	-0.024	-0.005	0.169	-0.006
Number of grades failed AND LNE score	0.088	-0.170	0.307	-0.223	0.252	-0.387

Note: Based on probit regressions in Table 3, including LNE scores

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

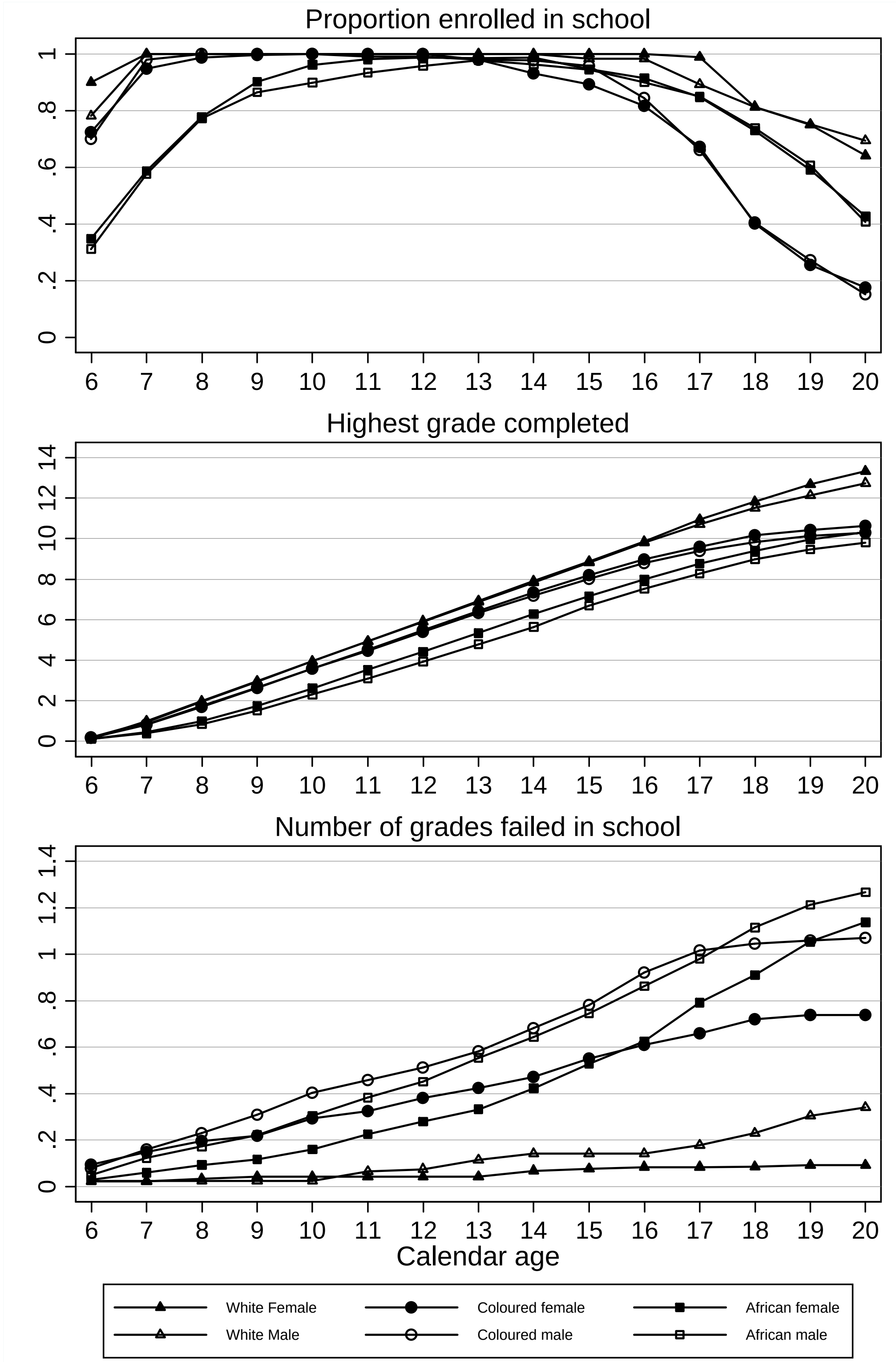


Figure 2. Impact of higher variance on probability of passing, enrollment, and effort

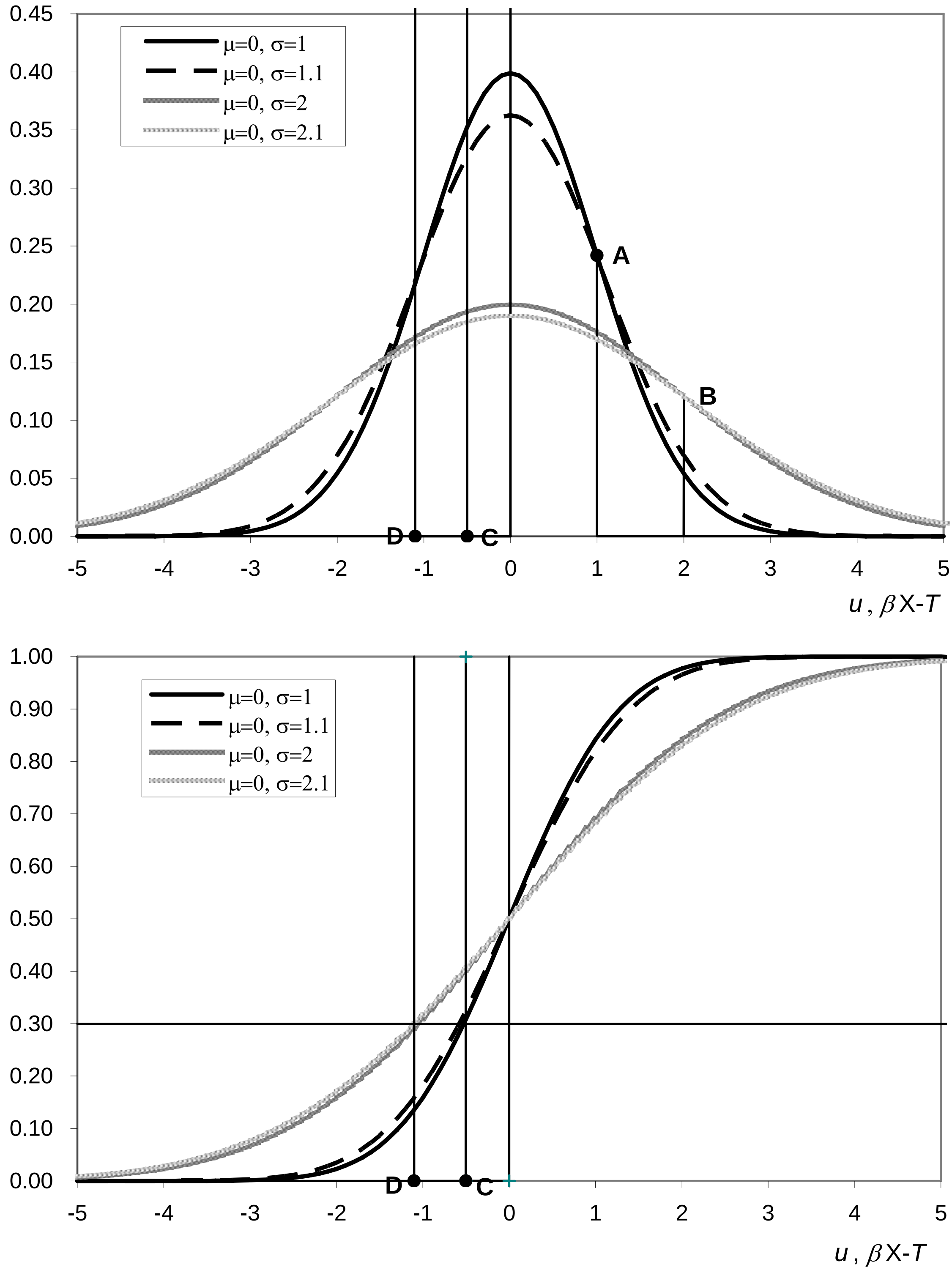
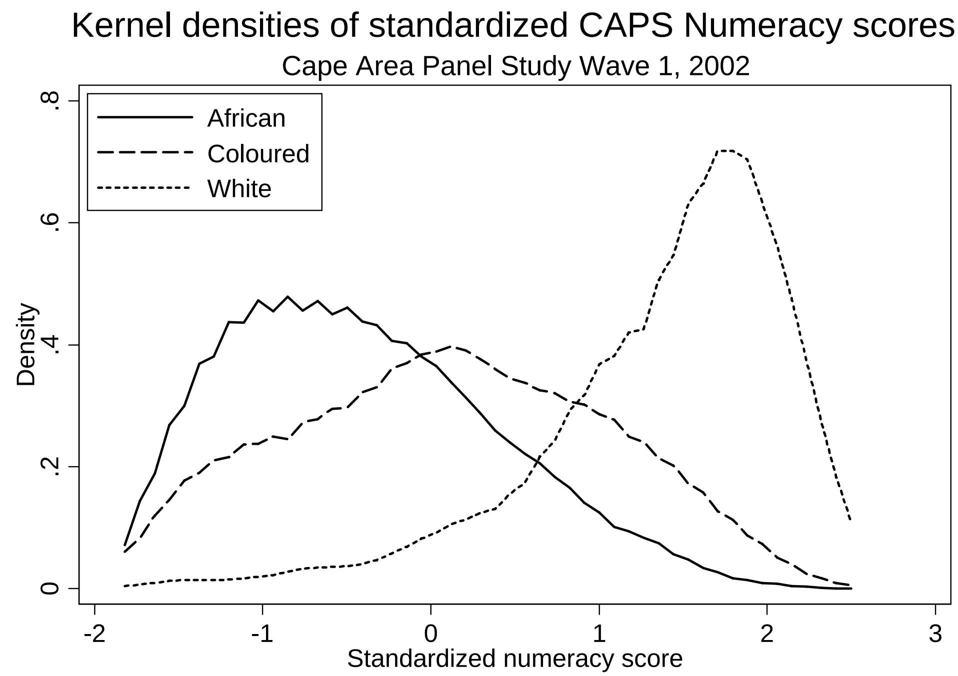


Figure 3. Kernel densities of CAPS numeracy and literacy scores



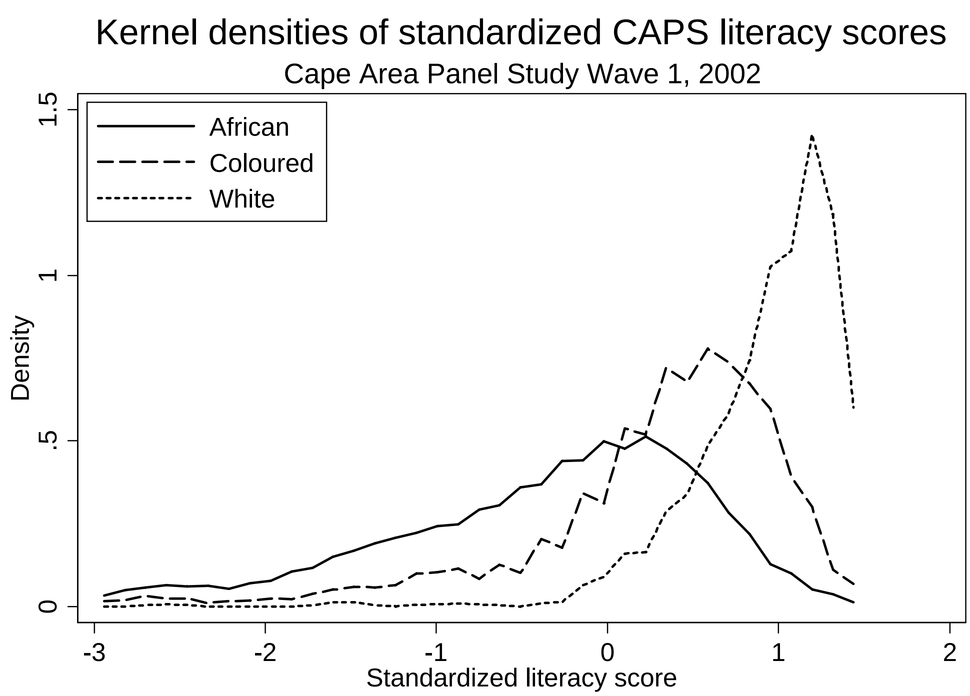


Figure 4. Kernel densities of log per capita household income, CAPS Wave 1

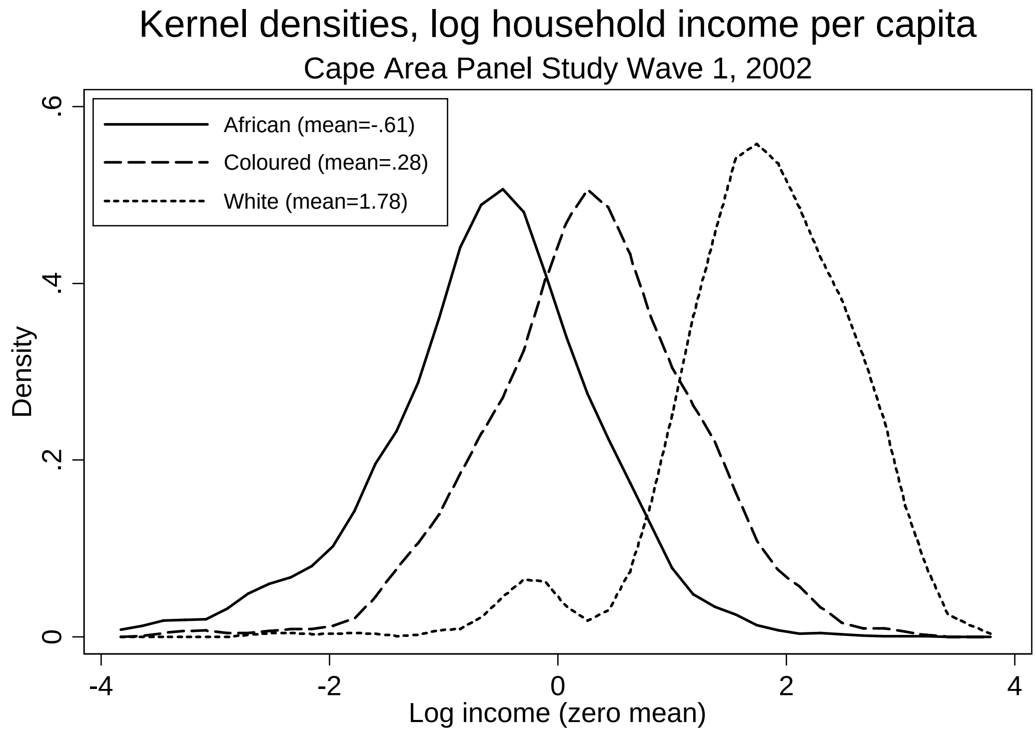
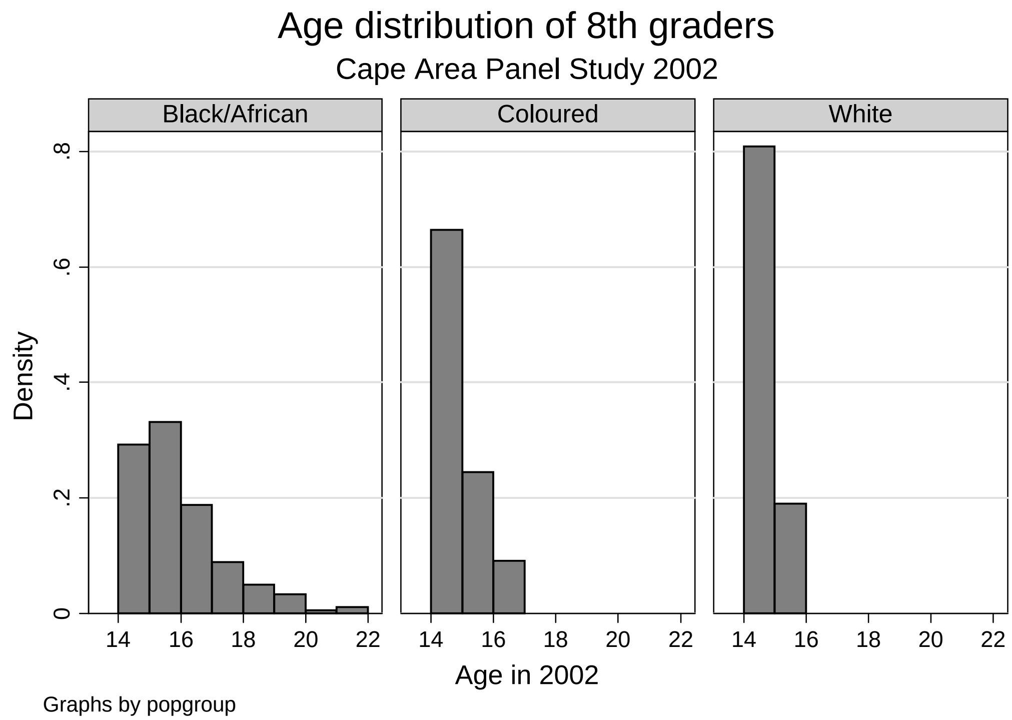


Figure 5. Age distribution of 8th and 9th graders, CAPS Wave 1, 2002



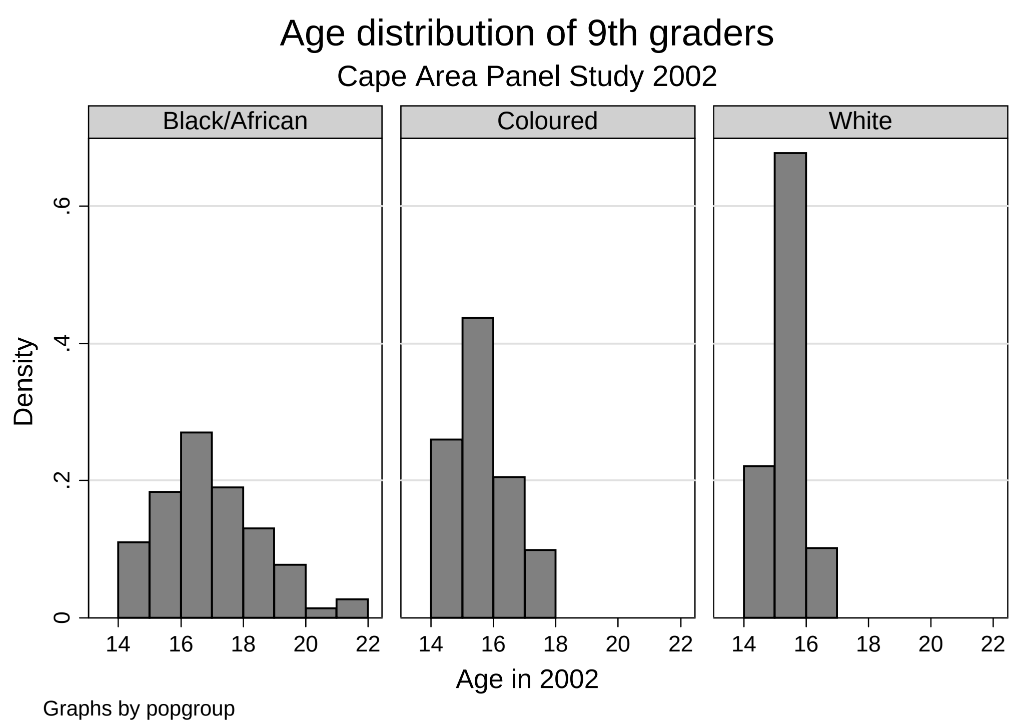


Figure 6. Predicted probability of advancing three grades by 2005 and being enrolled in 2004, by LNE score and per capita household income