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

David Lam[a,*], **Cally Ardington**[b], and **Murray Leibbrandt**[b]

[a]Department of Economics and Population Studies Center, University of Michigan, 426 Thompson Street, Ann Arbor, Michigan 48106, USA

[b]Southern Africa Labour and Development Research Unit, School of Economics, University of Cape Town, Cape Town, South Africa

Abstract

This paper analyzes the large racial differences in progress through secondary school in South Africa. Using recently collected longitudinal data we find that grade advancement is strongly associated with scores on a baseline literacy and numeracy test. In grades 8-11 the effect of these scores on grade progression is much stronger for white and coloured students than for African students, while there is no racial difference in the impact of the scores on passing the nationally standardized grade 12 matriculation exam. We develop a stochastic model of grade repetition that generates predictions consistent with these results. The model predicts that a larger stochastic component in the link between learning and measured performance will generate higher enrollment, higher failure rates, and a weaker link between ability and grade progression. The results suggest that grade progression in African schools is poorly linked to actual ability and learning. The results point to the importance of considering the stochastic component of grade repetition in analyzing school systems with high failure rates.

Keywords

education; grade repetition; South Africa

1. Introduction

Grade repetition is one of the most important problems in educational systems in many developing countries. In sub-Saharan Africa, where the problem is particularly severe, repetition rates are often 20% per grade (Lee et al. 2005), contributing both to low average levels of schooling and high schooling inequality. In spite of the wide recognition of the importance of grade repetition, research on the determinants of progress through school

[*]Corresponding author. davidl@umich.edu (D. Lam), Cally.ardington@uct.ac.za (C. Ardington), Murray.leibbrandt@uct.ac.za (M. Leibbrandt).

remains very limited. The goal of this paper is to advance our understanding of grade repetition by analyzing progress through secondary school in South Africa. More than a decade after the end of apartheid there continue to be large racial differences in schooling outcomes in South Africa. As we will show, grade repetition plays a key role in explaining these differences.

South Africa has almost universal primary school enrollment, with enrollment rates remaining high into the teenage years (Anderson et al. 2001). Ultimate schooling attainment is mostly determined between ages 14 and 22, 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. We use the Cape Area Panel Study (CAPS), a longitudinal survey of youth in Cape Town, to follow students through three years of secondary school. We find large racial differences in grade advancement – 82% of white students who were in grades 8 and 9 in 2002 successfully advanced three grades by 2005, compared to 34% of coloured students and only 27% of African students. While dropping out is one reason for these differences, we show that high rates of grade repetition play a fundamental role. While only 27% of African students in grades 8 and 9 in 2002 had advanced three grades by 2005, 67% were still enrolled in school.

The importance of grade repetition has been pointed out in a number of developing countries. Gomes-Neto and Hanushek (1994) documented repetition rates of 20-54% per grade in primary school in Brazil. They found that lower test scores were associated with increased probability of grade repetition, a result consistent with our results. Jacoby (1994) found that 21% of 7-12 year-olds had repeated at least one grade in Peru. He found that household income and assets reduce grade repetition, concluding that borrowing constraints play an important role. As pointed out by Lee et al. (2005), grade repetition is an even more serious problem in sub-Saharan Africa, with repetition rates of over 20% per grade in many countries. Although the importance of these high rates of grade repetition is widely recognized, research on grade repetition is limited. This is due in part to data limitations, with few data sets providing direct information on grade repetition. The CAPS data were collected with a strong focus on grade repetition, allowing us to get a clearer picture of this important component of schooling inequality in South Africa.

As a framework for understanding progress through school we develop a stochastic model of grade advancement. Performance in school in a given year depends on systematic components such as prior learning, student effort, and inputs from home and school, as well as a stochastic component that reflects imperfect links between actual learning and measured performance. We show that high variance in this stochastic component can generate an equilibrium characterized by high enrollment 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 prior learning 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 rich set of variables from the CAPS. These variables include previous school outcomes, scores on a baseline literacy and numeracy evaluation, and household variables such as income and parental schooling. Our empirical results are highly consistent with our theoretical model. While there is a strong impact of baseline test scores and household income on progress through grades 8-11, the effect is much weaker for African students than for coloured and white students. We interpret this as evidence that the African school environment does a poor job translating ability and resources into measured performance. Also, in line with our model, we find that African students are less likely to drop out of school than coloured students after failing a grade. As a strong test of our model,

we show that our results change systematically when we look at pass rates on the nationally standardized grade 12 matriculation exam. The impact of baseline test scores and income are as large for African students as coloured students in predicting pass rates on the grade 12 exam. This suggests that the weaker impact of baseline test scores and income for Africans in grades 8-11 is due to a poor system of evaluation in those grades.

2. Historical Background and Empirical Regularities

2.1. Education and apartheid

Although educational attainment in South Africa is relatively high compared to many African countries, cross-national standardized tests show South African students lagging behind students in other countries, including many African countries (van der Berg and Louw 2007). One obvious explanation for this poor performance is that it reflects a lingering legacy of the extreme inequality in education that existed under apartheid. The government ran separate school systems for different racial groups, with enormous differentials in funding levels and in the design of the curriculum (Fiske and Ladd 2004). Although government funding levels were equalized across schools after 1994, there continue to be large racial differences in progress through school and ultimate educational attainment (Bhorat and Oosthuizen 2008, van der Berg 2007).

One reason that equalization in access and government funding has not led to equalization of educational outcomes is that there continues to be large inequality in school resources. School fees, which vary enormously even in government-run schools, play an important role in this inequality. There has been some reduction in the inequality in pupil-teacher ratios that was shown by Case and Deaton (1999) to have an important impact on inequality in schooling outcomes in 1993. Due to large disparities in school fees, however, the equalization of government funding has not fully equalized pupil-teacher ratios and other school inputs (Fiske and Ladd 2004; Yamauchi 2005). Although there are greater possibilities to exercise school choice in the post-apartheid environment, most black students are still in schools with poor educational infrastructure.

Research using an education production function approach has analyzed the role of these input inequities on educational performance (Case and Deaton 1999; Crouch and Mabogoane 1998 2001; Van der Berg 2007; Bhorat and Oosthuisen 2008). The overriding conclusion is that a large part of student performance remains unexplained after controlling for infrastructure differences. This suggests that less quantifiable aspects of school quality such as school management and teacher quality may play an important role. Hoadley (2007) concludes that South African schools have large systemic problems and struggle to meet their educational mandates in the three core functions of teaching, learning and management. She documents high teacher absenteeism, especially in more poorly resourced schools. Even in schools in which there is not a culture of absenteeism, problems of crowded classrooms, ineffective administration, and limited resources lead to a chaotic school environment. A ministerial review of school governance concluded that many school management teams cannot fulfill the functions allocated to them (Department of Education 2004).

In addition to the problems disadvantaged schools face in providing quality classroom instruction, there is evidence that they also struggle to effectively evaluate student performance. A national study of teacher workloads found that teachers in disadvantaged schools had a difficult time meeting the continuous assessment goals of the Outcomes Based Education curriculum that began in 1995 (Chisholm et al. 2005). While teachers reported spending a great deal of time on assessment, they found the assessment guidelines overly

burdensome, repetitive, and unnecessary. The study suggests that teachers with large classes in poor schools have a difficult time effectively evaluating student performance.

A comparison of school-based continuous assessment marks with externally evaluated matriculation examination marks provides direct evidence of the poor quality of internal assessment in many schools (van der Berg and Shepherd 2008). Most schools were found to have average internal assessments that exceeded their average external matric results. This gap was greatest among those schools with the poorest matric results. These low-performing schools also had the highest variance in their internal assessments. While van der Berg and Shepherd did not look explicitly at race, they found that internal assessments were more accurate in schools classified as higher socio-economic status. Most African students attend the low socio-economic status schools shown to have poor internal assessment. In van der Berg and Shepherd's view, inaccurate assessment is indicative of teachers having inadequate subject knowledge and a poor understanding of the demands of the curriculum.

It is important to keep in mind that school characteristics are only part of the story. Another important long-run impact of apartheid is that it leaves black parents without the resources to create a favorable home environment for students. It is therefore important to incorporate household characteristics into studies of school outcomes. The CAPS data provide us with detailed information on young people and their households. Because it is a household survey rather than a school-based survey, we are able to follow young people over time, whether or not they remain in school. This permits us to study, for example, whether failing a grade leads students to drop out of school, and allows us to link baseline characteristics with later school outcomes.

A major focus of this paper is comparing schooling outcomes for African, coloured, and white youth. These three groups were treated very differently under apartheid. Whites had advantages in most areas, including significantly higher expenditures on schooling, privileged access to the labor market, unrestricted residential mobility, and better access to social services. Africans had the least access to services and the most restrictions on work and migration, with a large gap in school expenditures. The coloured population, which is heavily concentrated in the Western Cape (including Cape Town), occupied an intermediate status under apartheid, with higher expenditures on schooling, fewer restrictions on residential mobility, and better access to jobs than Africans. This history of racial inequity in education is more than a matter of historical interest. As we document below, there continue to be enormous racial differences in variables such as school fees (which translate into school resources), student-teacher ratios, and household income.

2.2. Data: The Cape Area Panel Study

This paper uses the Cape Area Panel Study (CAPS), a longitudinal survey in metropolitan Cape Town.[1] Wave 1, collected in 2002, included 4,752 young people aged 14-22. Cape Town has three predominant population groups – the distribution in the 2001 census was 48% coloured, 32% African/black, and 19% white. CAPS oversampled areas classified as predominantly African and white. Cape Town is the only major city in South Africa to have substantial numbers of white, coloured, and African residents, providing unique opportunities to study changes in inequality after the end of apartheid.

The Wave 1 young adult questionnaire, administered to up to three household members aged 14-22, covered a wide range of variables including schooling and work. It also included a literacy and numeracy evaluation (LNE) which features prominently below. We use 2005,

[1]Details about CAPS, a collaborative project of the University of Cape Town and the University of Michigan, are available in Lam et al. (2008) and on the CAPS web site, www.caps.uct.ac.za.

the timing of Wave 3, as the endpoint for the transitions we analyze. Data from Wave 4, collected in 2006, is used to fill in data for respondents who were not interviewed in Wave 3. Table 1 gives information on sample size and attrition for respondents who were in grades 8 to 12 in 2002, the sample used in the analysis below. There were 2,479 respondents in grades 8-12 in 2002, 47% of whom were African. The "weighted percent" column shows that when we adjust for oversampling Africans are 30% of those in grades 8-12. The white sample is considerably smaller, a result in part of a lower response rate.[2] Looking at the sample with data for 2005, the overall attrition rate was 12%, with significant differences across population groups. The African attrition rate is 13%, with most attrition due to migration back to the rural Eastern Cape province that is the main sending region for Africans in Cape Town. The coloured population has its roots primarily in Cape Town, a factor contributing to its lower 6% attrition. The 24% attrition for whites includes both migration out of Cape Town (including out of South Africa) and a significant number of refusals.

2.3. School enrollment, grade repetition, and work

This section provides an overview of key patterns in school enrollment, grade repetition, and labor force activity that form a backdrop for the school transitions we analyze below. Figure 1 shows three indicators of schooling at each age from 6 to 20 based on retrospective reports of CAPS respondents aged 20-22 in 2002. The top panel shows the proportion who were enrolled in primary or secondary school (through grade 12) at each age. Enrollment rates for all groups are close to or above 90% for all ages between 9 and 15, with female enrollment slightly higher than male enrollment for all three population groups until around age 18. Africans lag behind in starting school. 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 coloureds and whites. Coloured enrollment rates begin to fall above age 15, with Africans having higher enrollment rates than coloureds at all ages above 15. Enrollment rates for whites drop rapidly at age 18, a reflection of the fact that most whites complete grade 12 by that age.[3]

The second panel of Figure 1 shows the number of grades completed at each age. Whites advance almost one grade per year, reaching a mean of about 8 grades by age 14. Although coloureds start school at about the same age as whites, and have similar enrollment rates, they lag behind whites in grade advancement from an early age. By age 14 coloured females were about 0.5 grades behind white females, with a similar gap for males. Africans start school later and advance more slowly. By age 14 grade attainment was 5.8 grades for African males, two full grades behind white males. Because of high enrollment rates for Africans in the late teens, Africans almost catch up with coloured grade attainment by age 20. The figure also shows a female advantage in grade attainment in all three groups. As pointed out by Anderson et al. (2001), girls move through school faster than boys, with female schooling exceeding male schooling by about one full grade among recent African cohorts who have finished schooling.

A valuable feature of CAPS is that it provides direct measures of grade repetition. For each grade respondents were asked whether they passed, failed, or dropped out before completing the grade. The bottom panel of Figure 1 shows the cumulative number of primary and

---

secondary grades failed at each age. Coloured and African students fail at a much higher rate than whites, with higher failure rates for males. African and coloured males fail an average of one grade by age 17. 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 faced by young people. Decisions about whether to stay in school and how hard to work in school will be affected by the opportunity cost of 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 2002 survey. Work is defined broadly, including work during school vacations, so work does not necessarily directly compete with school. There are enormous differences in the work experience across racial groups. At age 17 over half of white males and females report working in the last year, compared to 1% of African females and 7% of African males. Coloured youth are in between, with 26% of 17 year-old males and females having worked in the last year. At age 22 only 24% of African females and 35% of African males report working in the last year, compared to over 75% of the other four gender/race groups.

Summarizing the patterns in Figure 1 and Table 2, African teenagers in Cape Town have high rates of school enrollment, high rates of grade repetition, and low rates of employment. These patterns are similar to those for African youth in all of South Africa (Anderson et al. 2001). Limited labor market opportunities, driven in part by spatial segregation that is a legacy of apartheid, are presumably important in explaining both low employment and high enrollment. Coloured youth have significantly higher employment rates than African youth, a reflection of both closer geographic proximity to jobs and the legacy of 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 the highest rates of employment along with the highest levels of school enrollment and schooling attainment, an indication that work and school in the teenage years are not entirely incompatible.

3. A Stochastic Model of Grade Repetition

While we will not develop a complete theoretical model of school enrollment and progress through school, in this section we discuss a number of important theoretical issues that guide our empirical analysis. We pay close attention to the combination of high enrollment rates and high rates of grade repetition documented in Figure 1. At the simplest level, advancing through secondary school requires that a student achieves some level of learning sufficient for grade promotion, that the student's teachers and school correctly recognize that learning, and that the student chooses to continue in school as grades are either passed or failed.

We assume enrollment decisions are based on a calculation by the student (and family) that the expected benefits of enrollment exceed the direct costs and opportunity costs in a given year. Credit constraints may complicate this decision, with some families unable to pay the cost of schooling even when the present value of expected benefits exceeds the annual cost. Time devoted to schooling may also be affected by the opportunity cost of student's time, especially for poor, credit-constrained households. Household income may therefore be an important determinant of both enrollment and advancement. Household income may also be important because it proxies for a wide range of inputs that may have affected learning throughout the child's life. Theoretical models exploring these issues in both static and

dynamic settings have been developed in previous literature, and many of the important points are well known.

We focus our attention on a theoretical issue that has received much less attention but which we believe plays an important role in environments with high rates of grade repetition. This is the imperfect evaluation of student performance. In addition to the fact that crowded schools with poor infrastructure and weak administration do not teach well, they are also likely to do a worse job evaluating students than better-equipped and better managed schools. While students everywhere tend to rationalize failure to be 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 stochastic model of grade advancement. Suppose that students are evaluated at the end of the school year based on a final score S. One component of S is the students' actual knowledge at year's end, which we characterize by a learning production function $K=F(X)$, where $\mathbf{X}$ is a vector of inputs such as prior knowledge, effort, school inputs, and family background characteristics such as parental schooling and household income. The score for student i also includes a stochastic component u_i reflecting discrepancies between knowledge and evaluation. In the most literal sense these include errors in marking exams. More broadly they include problems in the school environment that cause learning to be unrewarded in grade promotion decisions. For example, weak teachers in bad schools may teach and test in such a disorganized way that mastery of course material has little impact on final evaluations. Our review above drew attention to frequently cited problems in South African schools such as overcrowding, teacher absenteeism, and disorganized school administration. These contribute to an environment in which there is a weak link between actual learning and measured performance.

Assume that we can summarize this environment with a linear model

$$S_i = \beta' \mathbf{X}_i + u_i \tag{1}$$

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 $\mathbf{X}$. We assume there are a large number of independent components in u, making it reasonable to assume that it is normally distributed, $u \sim N(0,\sigma)$. 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

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

where Φ is the cumulative of the standard normal distribution.

3.1. The effect of characteristics on passing

We can use (2) to analyze the impact on passing of some characteristic which is a component of $\mathbf{X}$, such as previously acquired human capital, parental education, or household income. Denote this variable by X_1, and its corresponding coefficient in (1) by β_1. 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. We differentiate (2) to get

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

(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. It is clear from (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 σ. For those near the passing threshold, a higher variance in the random component of the score reduces the marginal payoff to an extra point, and thus reduces the impact of characteristics. Defining $g_i = T - \beta'\mathbf{X}_i$ as the gap between the deterministic component and the threshold, and taking the derivative of Equation (3) with respect to σ,

$$\frac{\partial^2 P}{\partial X_{1i}\partial\sigma} = \left(\frac{\beta_1}{\sigma}\right) f\left(g_i\right)\left|\frac{g_i^2}{\sigma^2} - 1\right|.$$

(4)

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

Consider two identical students in two different schooling systems that have the same β_1 coefficient but have different stochastic variance, $\sigma_2 = 2\sigma_1$. Suppose both students are exactly at the passing threshold based on the deterministic component, $T = \beta'\mathbf{X}_2 = \beta'\mathbf{X}_1$, implying that they will pass with a positive draw of u and fail with a negative draw. Looking at (4), the marginal effect of one additional point on the probability of passing is twice as high in the low-variance regime. An increase in the value of some characteristic that causes a β_1 point increase in the deterministic component will have twice as large an impact on the probability of passing in the low-variance environment, evaluated for a student near the passing threshold.

Equation (3) reminds us of an econometric point that is important in our empirical analysis below. When 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 different probit coefficients for two different groups we generally cannot distinguish between differences in the marginal impact of the characteristic on learning (differences in β) and differences in the variance in the process that determines promotion (differences in σ). However, if the school environment provides a situation in which we expect smaller differences in σ between groups, we can use this to make some headway in distinguishing between differences in β and σ. As explained below, a nationally standardized exam at the end of secondary school provides such a situation.

3.2. Who goes to school?

Assuming that school enrollment is a voluntary decision by children and/or their parents, those who enroll in a given year will be those for whom the expected benefits exceed the expected costs. To simplify, suppose that attending school in a given year has zero payoff if the student does not pass that grade. If enrollment requires no out-of-pocket expenses and has no opportunity cost, then every student should enroll since every student has some probability of passing. Even those with low deterministic components of their final score have some probability of getting a lucky draw from u and receiving a passing score. More realistically, there are both direct costs and opportunity costs to being in school. There will therefore be some threshold probability of passing required for students to enroll in a given

year. To see how this probability is affected by the variance of the stochastic component, we take the derivative of (1) with respect to σ :

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

(5)

Equation (5) tells us that the probability of passing increases with σ for those who would fail based on the deterministic component, and decreases for those who would otherwise pass. As σ increases the expected probability of passing is increasingly determined by the stochastic component. Consider two groups of students, a low-skilled group for whom $T > \beta'\mathbf{X}_1$ and a high-skilled group for whom $T < \beta'\mathbf{X}_2$. If the opportunity cost is the same for the two groups 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 σ will tend to decrease the enrollment of high-skilled students while increasing 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*.

It is easy to simulate examples to illustrate the model's predictions. Suppose, for example, that the passing threshold is 50 points. There are two groups of students, each with a mean of 55 and a standard deviation of 10 in the absence of the stochastic component. The stochastic component has mean zero for both groups, with $\sigma=10$ for Group A and $\sigma=20$ for Group B. If students enroll who have an expected probability of passing of at least 30%, then we will observe the following: the enrollment rate is 85% in Group A and 94% in Group B; the passing rate among those enrolled is 72% in Group A and 61% in Group B; the average impact of one additional point on the probability of passing is 54% larger in Group A than Group B. In other words, we generate differences similar to those observed between African and coloured students in Cape Town – the group with the larger stochastic component has higher enrollment, lower pass rates, and a lower impact of characteristics on passing.

3.3. Effects over multiple years

Following students over multiple grades, we can generalize (1) by adding a subscript t and making assumptions about the correlation of stochastic terms across years. The simplest case is to assume that u_{t+k} is uncorrelated with u_t for all k, an assumption that fits our characterization of the idiosyncratic nature of the stochastic term. The probability of passing all years from year 1 to year n is $P_{i,1n} = \prod_{t=1}^{n} P_{i,t} = \prod_{t=1}^{n} P\left(S_{i,t} > T_t\right)$. Consider the simple case in which $\mathbf{X}$, β, σ, and T are the same every year, so that students get the same score each year in the absence of the stochastic term. If the stochastic terms are uncorrelated across years then the probability of passing is identical every year, $P_1 = P_t = P_n$. To analyze the impact of characteristics on the probability of passing n grades, it is helpful to take logs and look at the proportional impact. Taking the derivative of $\ln(P_{i,1n})$ with respect to a characteristic X_1, and using the result from (3), we get

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

(6)

where P_t is the probability of passing in any single year.

Equation (6) shows that the proportional impact of characteristics on the probability of passing increases as we look over more grades. If one IQ point gives a student a 2% higher probability of passing one grade, then it gives her a 10% higher probability of passing five consecutive grades.[4] While (6) is derived assuming that every grade is identical, the result is quite general as long as part of the stochastic component is uncorrelated across years. The stochastic component introduces noise into each year's results, causing some weak students to pass over better students and weakening the link between ability (for example) and scores. Over multiple years the better students pull ahead as the systematic component dominates the uncorrelated stochastic component. The uncorrelated components, which in our model represent noise in the link between learning and evaluation, become less important when we look across more years.

3.4. Impact of failing on enrollment and future success

Another implication of our model is that the impact of failing grades on future enrollment depends on the magnitude of the stochastic component. If students are uncertain about their ability and likelihood of future success in school, then each year's scores (and promotion decisions) are important signals about that ability. A larger stochastic component implies that grade promotion is a noisier signal about the student's ability and future probability of success. We expect, then, that past failure will be a weaker predictor of future enrollment in a regime with higher variance. Past failure 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.

3.5. Externally evaluated standardized exams

An important feature of the South African school system is the nationally standardized, externally evaluated matriculation exam given at the end of grade 12. The national Education Department oversees a process in which exam papers are pooled and graded by external evaluators. Performance on the exam has important consequences for both students and schools, with extensive media coverage of matric pass rates when they are announced each December. 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 there are important differences between passing grade 12 and passing grades 8-11. Since the standardization and external evaluation should lead to both a reduction in variance and smaller differences in the variance of the stochastic component across racial groups, we expect there to be a larger impact of characteristics on pass rates and smaller racial differences in the impact of characteristics.

While many factors affect school outcomes and enrollment decisions, this stochastic model of grade repetition captures some important features of the South African school environment. Most importantly, the model easily generates an equilibrium which has both high enrollment rates and persistently high rates of grade repetition. The model has a number of empirical implications that can be analyzed using CAPS. First, 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 with 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

---

[4]More precisely, a .02 log probability difference in 1 year implies a .10 log probability difference in 5 years.

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.

4. Empirical Evidence

In this section we analyze empirical evidence on the determinants of progress through secondary school. We focus in particular on the extent to which empirical evidence is consistent with our stochastic model of grade repetition. Before presenting regression results we begin with a descriptive overview of grade progression for our sample of 8th and 9th graders in 2002.

4.1. Grade Progression between 2002 and 2005

Data from CAPS Waves 3 and 4 allow us to follow the progress of students from 2002 to 2005. Table 3 shows the status in 2005 of those where in grade 8 and 9 in 2002. While 93% of whites who were in grade 8 in 2002 advanced to grade 11 or 12 by 2005, the experience of African and coloured youth is very different. Among Africans who were in grade 8 in 2002, only 36% had reached grade 11. About the same percentage, 37%, were in grade 10 (two grades in three years). About 18% 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 than Africans to have advanced three grades. About 46% were in grade 11 or 12, with 13% in grade 10. A higher percentage of coloured youth appear to have dropped out to work, with 13% in the "not enrolled/working" category. African youth are much more likely to stay in school than coloured youth, in spite of their higher rates of grade repetition. About 82% of Africans who were in grade 8 in 2002 were still enrolled in school in 2005, compared to 64% of coloured students.

We see similar patterns for those who were in grade 9 in 2002. Among whites, 85% reached grade 12 by 2005. This compares to 29% for Africans and 42% for coloureds. About 29% of Africans were in grade 11 and 11% were only in grade 10. About 69% of Africans who were in grade 9 in 2002 were still enrolled in 2005, compared to 61% of coloured students. The patterns in Table 3 illustrate several predictions of our model. Coloured students have access to schools that are higher quality than traditionally African schools. Coloured students are more likely than African students to make normal progress and to drop out, with African students having both higher enrollment and higher failure rates. These issues are explored in greater detail below.

4.2. Characteristics affecting progress through school

In this section we provide an overview of some of the individual, household, and community characteristics we will use in our regressions. One interesting feature of CAPS is the literacy and numeracy evaluation (LNE) that was administered to all youth respondents in Wave 1. This was a self-administered 45-question test that took about 20 minutes to complete. Respondents could take the test in 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 African respondents, 43% of coloured respondents, and 64% of white respondents. In interpreting the results it is important to keep in mind that most white and coloured students took the test in their first language, while Africans took the test in a second language. It must also be noted, however, that English is the official language of instruction in African schools and is used for many tests such as the grade 12 matriculation exam. We use the LNE scores as a measure of cumulative learning at the time of the 2002 interview. Performance on the

test reflects a combination of many factors, including innate ability, home environment, and the quantity and quality of schooling to that point.

Figure 2 presents kernel density estimates of the distribution of the combined literacy and numeracy scores for each population group, using the sample of those enrolled in grade 8-12 in 2002 (the score is standardized to zero mean and unit variance for the full sample of 14-22 year-olds). Racial differences in test scores are striking. There is only a small area of overlap between the test scores of Africans and whites, with a much higher variance among Africans. The distribution of scores for coloureds sits between, with considerable overlap with both the white and African distributions. The mean standardized score is −0.6 for Africans, 0.01 for coloureds, and 1.14 for whites, implying a 1.7 standard deviation gap between whites and Africans. The standard deviation of African scores is 60% larger than the standard deviation of white scores.

Another key variable in our regressions is the log of per capita household income in 2002, as reported by an adult respondent in the Wave 1 household questionnaire. Figure 3 plots kernel densities for each population group, standardized to the overall mean. Once again we see large racial differences, with a difference in mean log income between whites and Africans of almost 2.5 (implying that white youth lived in homes with 10 times higher per capita income than Africans). As with the test scores, a striking feature is the very small range in which the African and white income distributions overlap, with the coloured distribution sitting in between.

An additional factor to consider in explaining school progress for 8th and 9th 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 and 9 there is considerable variation in the age of students. Figure 4 shows the age distribution for 9th graders in 2002. White 9th graders are concentrated at age 15, with only about 15% at age 16. By contrast, the modal age of African 8th graders is 16, with a wide distribution ranging between ages 14 and 22. Roughly 25% of African 9th graders are age 18 or older.

Table 4 presents descriptive statistics for students enrolled in grades 8 or 9 in 2002. The first row shows the dependent variable in our first regressions, an indicator of whether the student advanced three grades by 2005. This variable equals 0 for any other outcome, including dropping out by 2005 or being in a grade below the target grade in 2005. The percentage advancing three grades varies enormously by race: 27% for Africans, 34% for coloureds, and 82% for whites. Table 4 also shows the percentage enrolled in school in 2003, 2004, and 2005. Looking at enrollment in 2004, the outcome in our second set of regressions, there is large variation across racial groups. About 96% of whites were enrolled in 2004, compared to 82% of Africans and 69% of coloureds. Table 4 presents three measures of grade failure. The number of grades failed by 2002, which we use as an independent variable in our first regressions, varies from 0.8 for Africans to 0.6 for coloureds and 0.2 for whites. As shown in the next row, 52% of Africans failed at least one grade by 2002. The percentage who failed their grade in 2002, which we use in our regressions analyzing 2004 enrollment, varies from 17% for Africans to 2% for whites.

Table 4 includes means of several household characteristics that will be included in the regressions. The large differences in the log of per capita household income were already noted. The mothers and fathers of African youth have 4-5 years less schooling than the parents of white youth, with father's schooling missing for about 40% of Africans[5]. We also include in our regressions the age-sex-specific unemployment rate for individuals with less

than 12 years of schooling in the census sub-place. This varies from 77% for Africans to 26% for whites.

Table 4 also presents information about school characteristics. About 75% of Africans attend schools that were classified as African schools (Dept. of Education and Training) under apartheid; 13% attend formerly coloured (House of Representatives) schools; 3% attend formerly white (House of Assembly) schools; 9% attend schools that were created since 1994 and thus have no "former department" classification.[6] 86% of coloured students are in formerly coloured schools and 91% of white students are in formerly white schools. The school expenditure variable shows the large differences in school fees. African students paid an average of 326 rands per year (roughly 32 dollars), compared to R731 for coloureds and R5,817 for whites. Since these fees are used to hire extra teachers, the differences in fees translate into the differences in mean pupil-teacher ratios shown in the final row of Table 4 – 32.4 for Africans versus 24.0 for whites.

4.3. Probit regression for progress through school

This section presents probit regressions in which our dependent variable indicates progress through school between 2002 and 2005. One of our key empirical questions is whether there are racial differences in the impact of individual and household characteristics on grade advancement. Given the differences in school environment discussed above, we hypothesize that Africans have lower β coefficients in the learning production function and/or a higher σ for the stochastic component, both implying that Africans will have lower estimated probit coefficients (β / σ) in a regression of school advancement on characteristics. 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 Africans, coloureds, and whites. For each coefficient and each pairwise combination of races we test for equality of the probit coefficients. While it is impossible to distinguish between differences in β and σ from the probit regressions alone, we will argue below that restrictions imposed by the standardized grade 12 matriculation exam help us identify the separate contributions of these two components.

Table 5 presents the first set of probits, which analyze the probability that those enrolled in grade 8 and 9 in 2002 advanced three grades by 2005. Columns 1-3 present the coefficients, columns 4-6 present tests of equality of coefficients, and columns 7-9 present marginal effects evaluated at a common set of characteristics across samples.[7] We estimate large effects of the LNE score and the number of previous grades failed, demonstrating the importance of prior learning and school performance. This is consistent with the results of Gomes-Neto and Hanushek (1994), who found that test scores were an important predictor of grade repetition in Brazil. The number of previous grades failed has a much less negative effect on grade advancement for Africans than for coloureds and whites. At the assumed baseline characteristics, having failed one additional grade by 2002 is associated with a 7 percentage point decline in the probability of advancing 3 grades for Africans, compared to a 24 percentage point decline for coloureds.

The LNE score also has a smaller positive effect for Africans. The probit coefficient is 0.31 for Africans, compared to 0.74 for coloureds and 0.87 for whites. Referring back to our

---

theoretical model, this lower coefficient for Africans could result from either a smaller coefficient on LNE scores in the learning production function (β) or from a higher variance in the stochastic component of measured performance (σ), or some combination of the two. Looking at the marginal effect, a one standard deviation increase in the LNE score is associated with a 12 percentage point increase in the probability of advancing three grades for Africans, compared to 26 and 23 percentage point increases for coloureds and whites, respectively.[8]

The impact of log per capita household income is not statistically significantly for Africans, but is strongly positive for coloureds and whites. This may seem surprising, since we might expect large effects of income over the income range in the African sample. The poorest Africans are in deep poverty, while the upper tail has incomes that should make it easier to keep children in school and support grade progression. This differs from Jacoby's (1994) results for Peru, where income was an important predictor of grade repetition. Our interpretation of the 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 measured learning. The low impact of income for Africans is unlikely to result from greater measurement error in income for Africans since we will see below that income is a strong predictor of passing the grade 12 matric exam for Africans.

In Columns 4-6 of Table 5 we test for equality of coefficients between pairs of racial groups. We can reject the hypothesis that Africans and coloureds have equal coefficients on previous grades failed and LNE scores. 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 these same variables, in spite of large differences in the point estimates. We also cannot reject equality of the coefficients on income for any pairwise comparison, although both the coloured and white coefficients on income are more than double the African coefficient.

Looking at other variables in our probit in Table 5, we find no significant differences in grade advancement between males and females. This is consistent with the patterns shown in Figure 1 and in other research showing that there is no female disadvantage in schooling outcomes in South Africa, at least through secondary school. Parental schooling has surprisingly weak effects on grade advancement, with a statistically significant coefficient estimated only for mother's schooling in the coloured regression. This is surprising given the high variance in parental schooling and the extensive research that finds strong effects of parental schooling on children's schooling. For Africans the effect of father's schooling is significantly positive if we exclude income and prior performance (results not shown). For coloured students we estimate a significant positive effect of both parents' schooling when we exclude income and prior performance.

The neighborhood unemployment rate is not significant for any group. We include it to capture two possible effects. On the one hand, the opportunity cost of time may affect effort in school or the probability of dropping out. On the other hand, better employment prospects might motivate youth to stay in school and study harder. These effects may be cancelling out, although it is also possible that census subplace 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

---

[8]As dicussed in Section 5.7 below, these results are robust to a number of alternative specifications, including the use of quadratics in LNE scores and the use heteroskedastic probits.

may not be sufficient geographical variation in job opportunities within racial groups to identify an effect.

Looking at the predicted probabilities evaluated at a common set of characteristics, it is striking that the predicted probability of passing is highest for Africans. This result is very robust to the choice of baseline characteristics. When a single regression is estimated for the pooled sample and dummy variables are included for white and coloured (not shown), the white and coloured coefficients become negative when the regression includes the variables shown in Table 5. While it is important to keep in mind the small overlap in the distributions of African and white income and test scores, the results suggest that the large racial gap in progress through secondary school can be statistically accounted for by a combination of initial human capital (previous grades failed and LNE scores) and family background (income and parental schooling). These issues are explored in more detail in Ardington et al. (2009).

4.4. Regressions for school enrollment

Table 6 presents regressions in which the dependent variable is school enrollment in 2004, continuing to use the sample of respondents who were enrolled in grade 8 or 9 in 2002. We include a dummy variable for whether the respondent failed their grade in 2002 in order to see whether students drop out or return to school after failing. Other variables 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.

As in Table 5, we estimate negative effects of prior grades failed and positive effects of LNE scores. The point estimates are larger in magnitude for coloureds than for Africans, though the difference is only marginally significant for the grades failed variable. A one standard deviation increase in LNE score is associated with a 3.1 percentage point increase in enrollment probability for Africans and a 7.1 percentage point increase for coloureds. The estimated effect of income is statistically insignificant for Africans but strongly positive for coloureds, with the difference in coefficients statistically significant. While it is surprising that income does not affect African enrollment, we interpret it as indicating that the combination of low opportunity cost, high returns to schooling, and imperfect evaluation make the benefits of being enrolled sufficient to overcome direct costs of fees and uniforms. Failing the grade in 2002 has a negative effect on 2004 enrollment for both Africans and coloureds, but the effect is much greater for coloureds and we strongly reject equality of the coefficients. Failing in 2002 reduces the probability of enrollment in 2004 by 52 percentage points for coloureds, compared to 21 percentage points for Africans. This is consistent with our interpretation of the response of Africans and coloureds to differences in the school environment. Failing a grade is a weaker predictor of future success for Africans than for coloureds. Consistent with this, Africans are less likely to drop out if they fail their grade in 2002.

4.5. Grade 12 matriculation exam

The grade 12 matriculation exam provides an interesting comparison to grade advancement from grades 8 to 11. The matric exam is nationally standardized, externally evaluated, and is explicitly designed to test material taught in the secondary school curriculum. Prior to the grade 12 exam the decision about whether to pass a student is made at the school level, based on a combination of graded material during the school year, end-of-year exams, and subjective evaluation by teachers. Given the structure of the matriculation exam, matric pass rates should have a smaller stochastic component than the school-level pass decisions for grades 8-11, and the stochastic component of matric pass rates should be similar across racial groups. We should therefore find that the impact of prior learning on the probability of

passing the matric exam is larger and more equal across racial groups than was the impact of prior learning on grade 8-11 pass rates. In this section we present regressions in which the outcome is passing the grade 12 matriculation exam. The sample is all CAPS respondents who were enrolled in grade 12 in 2002, 2003, or 2004 and reported matriculation exam results. We also present separate regressions for passing grade 9, 10, and 11, using the sample of students who were enrolled in these grades in 2002, 2003, or 2004.

Table 7 presents pass rates and mean characteristics for the samples for each grade. Row 1 gives the mean pass rate the first time students took the matric exam. About 78% of Africans passed the exam on their first attempt, compared to 90% of coloureds and 99.6% of whites. Given the almost universal pass rate for whites we exclude them from our regressions. The second row shows the percentage who passed "with exemption," a higher pass that qualifies students for university admission. Only 18% of Africans passed with exemption, compared to 23% of coloureds and 59% of whites. The third row shows that 11% of African students took the exam more than once between 2002 and 2004. We will include these multiple attempts in our regression, correcting the standard errors for clustering at the individual level. Grade 12 students are, not surprisingly, a selective sample of students. Comparing 12th graders in the top panel with 9th graders in the bottom panel, the mean LNE score of 12th graders is about half a standard deviation above the mean for 9th graders for African and coloured students.

Table 7 also documents the large differences in pass rates across grades. Pass rates are low in grade 11, with only 64% of African students and 79% of coloured students passing. This is consistent with the widely held view that teachers and school administrators hold back grade 11 students who they feel are not ready to pass the matric exam, motivated in part by a desire to increase the school's pass rate. Interestingly, however, pass rates for Africans are equally low at grade 10 as grade 11, and for coloureds the 69% pass rate in grade 10 is the lowest of any grade. Grade 9 pass rates are higher, at about 82% for both African and coloured students.

Table 8 presents regressions in which the dependent variable is equal to 1 if the student passes a given grade. We are particularly interested in the results using the standardized grade 12 matriculation exam in the top panel. The most striking result of the grade 12 regression is that the coefficient on the LNE score is now slightly larger for Africans than it is for coloureds. This is in contrast to the lower impact of LNE scores for Africans that we saw on grade progression at grades 8-11 in Table 5. We also estimate a lower impact of LNE scores for Africans than coloureds in the separate regressions for grade 9, 10, and 11 in Table 8, although only the grade 10 estimates are statistically different at the 10% level.

Looking at marginal effects in Columns 4-5, the point estimates imply that a one standard deviation increase in LNE score is associated with a 12.7 percentage point higher probability of passing matric for Africans, compared to a 12.0 percentage point increase for coloureds. The effect of income for Africans is also larger and closer to the effect for coloureds in the matric regressions than it was in the regressions for advancing three grades or in the separate regressions for grades 9, 10, and 11. In our grade 12 results we continue to find a smaller impact of previous grades failed for Africans than for coloureds. This is expected, since our model implies that failing grades is less of a signal about prior learning for Africans. Indeed, the fact that previous failed grades have no significant effect on the probability of passing the matric exam for Africans, while it has a large highly significant negative effect for coloureds, is entirely consistent with our interpretation of the weak connection between learning and evaluation in African schools.

Comparing the results from the grade 12 regression with the results for grades 9-11, it is striking how much larger the impact of income and LNE scores are in grade 12 for Africans. The marginal impact of LNE scores on the probability of passing grade 12 is more than double its impact on passing grade 9 or 10 and 70% larger than its impact on passing grade 11. The marginal impact of income for Africans is small and often not significantly different from zero at grades 9, 10, and 11, but becomes strong and significant in grade 12. For coloureds the impact of LNE scores and income do not increase significantly at grade 12. All these patterns are consistent with a regime in which there is a weak link between learning and evaluation for Africans in grades prior to grade 12, with the situation suddenly changing at the standardized grade 12 exam.

Given our probits for advancing three grades in Table 5, the results for passing the matric exam in Table 8 are quite remarkable. While an extra point on the baseline LNE exam has less than half the impact on the probability of advancing three grades for African students compared to coloured students, an extra LNE point has roughly equal impact for Africans and coloureds on the probability of passing the matric exam. This provides strong support for our interpretation that the racial difference in the impact of LNE scores on grade advancement is due to a weaker link between learning and evaluation in African schools. Put another way, this suggests that it is a larger σ rather than smaller β s that cause Africans to have smaller probit coefficients in Table 5. If racial differences in the grade advancement regressions were due to an interaction between prior learning and school quality (causing lower β coefficients in the learning production function), then we should see the same kind of differences showing up in the matric regressions. The results suggest that initial human capital does translate into higher learning in both the African and coloured schools, but that this learning does not translate equally into grade advancement.

Table 8 also allows us to evaluate another prediction of our model – that the impact of characteristics on grade advancement increases as we look over a larger number of grades. As shown in Equation 6, the prediction is clearest for the proportional impact, since the absolute passing rate declines as we look across multiple grades. Combining the predicted probability of passing with the marginal effect of LNE scores evaluated at X_1 in Table 8, a one standard deviation increase in the score for Africans implies a 16% increase (0.127/0.792) in the probability of passing grade 12, a 12% increase at grade 11, an 8% increase at grade 10, and a 6% increase at grades 9. Using the marginal effects and predicted probability of passing in Table 5, a one standard deviation increase in the LNE score implies a 33% increase (0.115/0.343) in the probability of passing three grades. In other words, the proportional impact of the LNE score on passing three consecutive secondary grades is almost three times greater than the impact on passing grade 11 alone and four times greater than the impact on passing grade 10 alone.

Results for coloureds are similar: While a one standard deviation increase in the LNE score implies a 9% increase in the probability passing grade 9, a 26% increase in the probability of passing grade 10, and a 14% increase in the probability of passing grade 11, it implies a 71% increase in the probability of passing from grade 8 or 9 to grade 11 or 12. Similar results apply to the impact of household income. These results are entirely consistent with our stochastic model of grade repetition, suggesting that there is a random component of passing that is uncorrelated with ability and uncorrelated across years. The importance of this component declines when we look at passing over multiple grades, increasing the impact of characteristics such as the LNE score.

5. Alternative Explanations and Robustness Checks

In this section we consider alternative explanations of our results and present a variety of robustness checks. The results presented above are highly consistent with the predictions from our stochastic model of grade repetition. We do not want to overstate the extent to which this is the only explanation for observed patterns in grade repetition, however, and it is important to consider whether other factors can explain the patterns we observe. In this section we discuss other factors that may drive our results. While many of these may play an important role, none of them can fully explain the empirical regularities in the data. It is also important to consider whether the results are robust to a number of alternative specifications. We discuss a number of possible alternatives and show that our results are quite robust.

5.1. Selectivity effects

An important issue in analyzing determinants of matric pass rates is that students who take the matric exam are a select group that managed to reach grade 12. As seen in Table 7, students in grade 12 have higher baseline LNE scores, higher household income, and failed fewer grades than students in grades 9-11. This selectivity applies to both African and coloured students, with the differences in means between African and coloured students in grade 12 being similar to the differences in grades 9-11. Evidence on the role of selectivity is provided in Figure 5. Panel A presents lowess estimates of the proportion advancing three grades by LNE score, using the sample used for the regressions in Table 5. Panel C shows the distribution of test scores for this sample. As the figure makes clear, the gradient is flatter for Africans than for whites and coloureds across the entire distribution of LNE scores. In the area between −1 and 0 standard deviations, where the African and coloured distributions have significant overlap, the African gradient is much flatter than the coloured gradient.

Panel B in Figure 5 gives lowess estimates of the proportion passing matric by LNE scores, with the distribution of test scores for this sample shown in Panel D. While the distribution of LNE scores for the grade 12 sample is shifted to the right compared to the grade 8-9 sample, this is clearly not the reason why we find that the impact of LNE scores is similar for Africans and coloureds for passing grade 12 but is much flatter for Africans for advancing three grades. The slope differences in Panel A are observed over the entire distribution of test scores, while in Panel B the slopes are similar for Africans and coloureds over the entire distribution. These results suggest that the selectivity of grade 12 students is not the explanation for the different slopes we estimate between grade 9-11 on the one hand and grade 12 on the other.

5.2. Opportunity cost

It is almost surely the case that coloured youth face a higher opportunity cost of schooling than African youth, given differentials in wages and in the probability of finding a job. This surely plays a role in explaining why coloured youth are more likely than African youth to drop out of school. While differences in opportunity cost may help explain why coloured youth drop out after failing a grade, they cannot explain the two patterns shown in Figure 5 – the impact of cognitive skills on grade advancement is larger for coloured students than for African students, while the impact of cognitive skills on passing the matriculation exam is very similar for the two groups.

5.3. Measurement error

Measurement error in our independent variables may contribute to the differences in coefficients we estimate for different groups. The fact that the literacy and numeracy exam was taken in a second language by most Africans, for example, may mean that it is a noisier

measure of cognitive ability for Africans than for coloureds and whites. Household income may also be measured with more error for Africans due to more irregular forms of employment. While measurement error could help explain the smaller coefficients we estimate for Africans for advancing three years and in grades 9-11, this must be reconciled with the fact that we estimate roughly equal or even larger coefficients for Africans on the grade 12 exam. We therefore think that measurement error is unlikely to be an important factor driving our key results.

5.4. Social promotion

Pressure on teachers and schools to promote underperforming students could play a role in our results. If schools want to keep the failure rate down or decide that students must not fail a grade more than once, then some students may be passed with little connection to performance. To some extent this is consistent with our model, since it is one way to generate a weak link between performance and promotion. We think this plays only a minor role in driving our results, however. Looking at Figure 5, the phenomenon to be explained is not just that some low-skilled African students pass. It is also that a surprisingly high fraction of high-skilled African students fail. The gradient between LNE scores and passing is relatively flat over the entire distribution. It is also important to note that social promotion, to the extent it exists, by no means creates a situation in which everyone passes. As shown in Table 7, pass rates are under 70% for African and coloured students in grade 10. Looking at the relationship between passing and LNE scores, and looking at the extent to which passing predicts future matric performance, the 30% of African students who fail appear to be much more randomly selected than the 30% of coloured students who fail.

5.5. Other components of variance

While we have emphasized the component of the variance related to imperfect evaluation of student performance, there are many other components of the variance that could cause probit coefficients to differ across racial groups. If Africans have larger unmeasured differences in school quality, for example, this could lead to larger variance in learning that we fail to control for in our regressions. This would be empirically indistinguishable from a higher variance in the stochastic component of evaluation. A greater variance in unmeasured components of family background could have similar effects. Our argument on this issue is similar to the points made above. While other components of variance could explain why Africans have lower probit coefficients in grades 9-11, this must be reconciled with the fact that the African coefficients are equal or higher in grade 12. The reduction in variance related to evaluation is the most plausible explanation for the patterns we observe, suggesting that African schools have noisier evaluation in grades 9-11.

5.6. Including measures of school quality

Since we are arguing that African schools do a worse job of evaluating performance, it would be interesting to compare African and coloured students going to schools of similar quality. While we do have measures of school quality, including pupil-teacher ratios and school fees, there is so little variation in these variables in African schools that it is difficult to identify a group of "high-quality" African schools to compare with coloured schools. The school quality variables also have relatively little predictive value on our outcomes.[9] When we interact school fees with the LNE score we get a positive and significant coefficient for Africans, consistent with our argument that there is a larger impact of ability on passing in

---

[9]Our finding that measures such as pupil-teacher ratios are poor predictors of school outcomes is similar to much of the literature (Hanushek, 2006). As noted in Section 2.1, most observers agree that there is high variance in school quality in South Africa, including variance in the ability to do effective evaluation. Measures like pupil-teacher ratios, however, appear to be poor indicators of school quality.

higher quality schools.[10] However, given the weak predictive power of our school quality measures and the concern that school choice is endogenous, we do not put much emphasis on these results.

Another approach is to look at African students attending coloured schools. Unfortunately, as shown in Table 4, few African students attend coloured schools. If we re-estimate the probit in Table 5 using only the 50 African students in coloured schools, the coefficient on the LNE variable is 0.52 (standard error 0.30), 66% higher than the point estimate for Africans in Table 5. The coefficient on log income is 0.19 (standard error 0.26), double the point estimate in Table 5. While the small sample size makes these estimates imprecise, especially for income, they are evidence in favor of our argument that if African students attended better schools their characteristics would be better predictors of grade advancement.

5.7. Additional robustness checks

In this section we consider two alternative specifications. First, we add a quadratic term in the LNE score in the probits estimated in Table 5. This allows us to test whether the racial differences in slopes shown in Table 5 are the result of non-linearities in the relationship between LNE score and the probability of passing three grades. In the quadratic specification the marginal effects of the LNE score evaluated at zero (the mean) are 0.17 for Africans, 0.27 for coloureds, and 0.28 for whites. We see, then, that the lower slope for Africans is robust to the quadratic specification, as we would expect from Panel A in Figure 5.

We also estimate heteroskedastic probits, given that our model implies that the variance differs by race. If all groups are constrained to have the same βs, the heteroskedastic probit is very much consistent with our prediction – Africans have more than double the σ of coloureds and whites. This must be the case given the results in Table 5 – Africans have a higher β/σ, and the regression must put the difference on σ if the βs are constrained to be equal. If we allow all βs to differ by race for all variables (in addition to the σs), the heteroskedastic probit is not identified, since an infinite combination of βs and σs are consistent with the data. If we allow **β**s to differ by race for LNE and income (but not for other variables), we cannot reject that the **β**s are equal for Africans, whites, and coloureds and the estimated σ for Africans is double the σ for coloureds and whites (we strongly reject the hypothesis of equal σs). While these results are entirely consistent with our predictions, we do not focus on the heteroskedastic probit results since we think the identification of the separate β and σ components is tenuous.

6. Conclusions

Grade repetition is a fundamental feature of the secondary school environment in South Africa. Our Cape Area Panel Study data show that only 27% of African students who were in grade 8 or 9 in 2002 advanced three grades by 2005, even though 2/3 of them were still enrolled in school in 2005. Taking advantage of the panel data on school advancement and the rich set of baseline characteristics in CAPS, we show that baseline cognitive skills and household income are important determinants of progress through secondary school. We find that the effects of previous grades failed, literacy/numeracy scores, and household income are considerably smaller for African students than for coloured and white students, however, a result that is highly robust.

---

[10]We also estimate a positive interaction of LNE and school fees for coloureds, but the coefficient is not statistically significant. Interactions of LNE with pupil-teacher ratios have the predicted negative sign, but are not statistically significant.

The smaller effect of characteristics on grade advancement for African students is consistent with our stochastic model of grade advancement. We show that by increasing the variance in the stochastic component of grade advancement we generate an equilibrium that looks very much like African secondary schools in Cape Town – high enrollment rates, high rates of grade repetition, and a weak link between baseline characteristics and grade advancement. An important prediction of the model is that if we could equalize the stochastic component in evaluation across racial groups then we would get more equal effects of characteristics on the probability of passing.

Our regression results are highly consistent with our theoretical predictions. We find a strong effect of test scores and household income on the probability of grade advancement for all races, but the effect is significantly weaker for African students. While this could indicate 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 enrollment rates of 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. Strong evidence in support of our interpretation is provided by the fact that the impact of baseline test scores on the probability of passing the nationally standardized grade 12 matriculation examination is slightly higher for African than for coloured students, suggesting that the payoff to ability is equalized when the stochastic component of evaluation is equalized.

From a policy perspective the strong impact of household income and indicators of previous achievement such as test scores remind us of the importance of quality primary schooling and the disadvantages of growing up in poor households. The results highlight persistent racial differences in the schooling environment and the signals that this sends to students. In drawing attention to the differential translation of learning into grade advancement our model and empirical results highlight one particular serious policy challenge – the need to strengthen the link between assessment and actual learning. The fact that this link is so much stronger for African students when assessment is nationally standardized and external to the school points to serious weaknesses in the ability of these schools to adequately evaluate student ability and learning.

Our analysis has important implications beyond South Africa. Our results suggest that a school's ability to accurately assess performance and determine which students advance to higher grades is a critical and understudied dimension of school quality. This is obviously of greatest importance when significant fractions of students are held back, but it may be important in any system in which some students fail. With high rates of grade repetition throughout Latin American and sub-Saharan Africa, this dimension of schooling would seem to deserve closer study. The strong empirical support for our theoretical model suggests that there are high returns to thinking systematically about the impact of imperfect evaluation on enrollment, effort, and grade advancement in systems with high levels of grade repetition.

References

Anderson KG, Case A, Lam D. Causes and consequences of schooling outcomes in South Africa: Evidence from survey data. Social Dynamics. 2001; 27(1):1–23.

Ardington, C.; Lam, D.; Leibbrandt, M. Explaining the persistence of racial gaps in schooling in South Africa. Presented at the IUSSP International Population Conference; Marrakech, Morocco. September. 2009

Bhorat H, Oosthuizen M. Determinants of grade 12 pass rates in the post-apartheid South African schooling system. Journal of African Economies. 2008; 18(4):634–666.

Cameron SV, Heckman JJ. The dynamics of educational attainment for black, Hispanic, and white males. Journal of Political Economy. 2001; 109(3):455–499.

Case A, Deaton A. School inputs and education outcomes in South Africa. Quarterly Journal of Economics August. 1999; 114(3):1047–84.

Chisholm, L.; Hoadley, U.; Kivilu, M.; Brookes, H.; Prinsloo, C.; Kgobe, A.; Mosia, D.; Narsee, H.; Rule, S. Educator workload in South Africa. Human Sciences Research Council; Pretoria: 2005.

Crouch L, Mabogoane T. When the residuals matter more than the coefficients: An educational perspective. Studies in Economics and Econometrics. 1998; 22(2)

Crouch L, Mabogoane T. 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. 2001; 27(1):60–78.

Department of Education. Report of the Ministerial Review Committee on School Governance. Department of Education; Pretoria: 2004. Review of school governance in South African public schools.

Fiske, E.; Ladd, H. Elusive equity: Education reform in post-apartheid. Brookings Institution; Washington, D.C.: 2004.

Gomes-Neto JB, Hanushek EA. Causes and consequences of grade Repetition: Evidence from Brazil. Economic Development and Cultural Change. 1994; 43(1):117–148.

Hanushek, EA. School resources. In: Hanusek, EA.; Welch, F., editors. Handbook of the Economics of Education. Elsevier; Amsterdam: 2006. p. 865-908.

Hoadley, U. The boundaries of care: education policy interventions for vulnerable children. Paper presented at Education & Poverty Reduction Strategies: Issues of Policy Coherence Conference; 21 - 23 February; Pretoria: 2007.

Jacoby H. Borrowing constraints and progress through school: Evidence from Peru. The Review of Economics and Statistics. 1994; 76(1):151–160.

Lam, D.; Ardington, C.; Branson, N.; Case, A.; Leibbrandt, M.; Menendez, A.; Seekings, J.; Sparks, M. The Cape Area Panel Study: Overview and technical documentation of Waves 1-2-3-4. The University of Cape Town; 2008.

Lee VE, Zuze TL, Ross KN. School effectiveness in 14 sub-Saharan African countries: Links with 6th graders' reading achievement. Studies in Educational Evaluation. 2005; 31:207–246.

van der Berg S. Apartheid's enduring legacy: Inequalities in education. Journal of African Economies. 2007; 16(5):849–880.

van der Berg, S.; Louw, M. Lessons learnt from SACMEQ II: South African student performance in regional context. University of Stellenbosch Economic Working Papers 16/07. 2007.

van der Berg, S.; Shepherd, D. "Signalling performance: An analysis of continuous assessment and matriculation examination marks in South African schools" Umalusi Council for Quality Assurance in General and Further Education and Training. Pretoria: 2008.

Yamauchi F. Race, equity and public schools in post-apartheid South Africa: Equal opportunity for all kids. Economics of Education Review. 2005; 24:213–33.
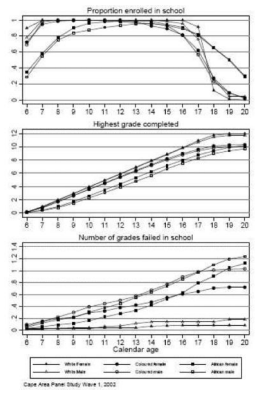
Figure 1.
Schooling experience from retrospective histories CAPS respondents 21-22. 2002
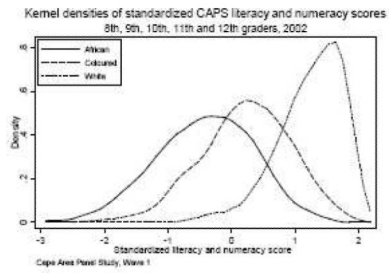
Figure 2.
Kernel densities for scores on CAPS Wave 1 Literacy and Numeracy Evaluation
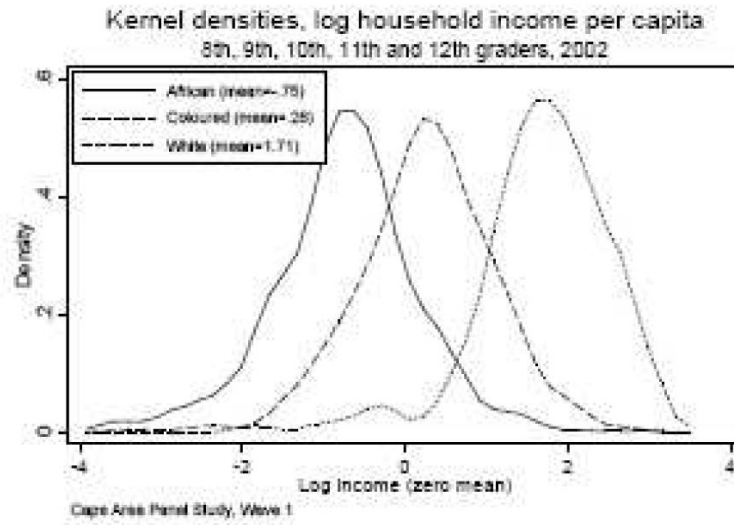
Figure 3.
Kernel densities of log per capita household. CAPS Wave 1

Age distribution of 9th graders

Cape Area Panel Study 2002
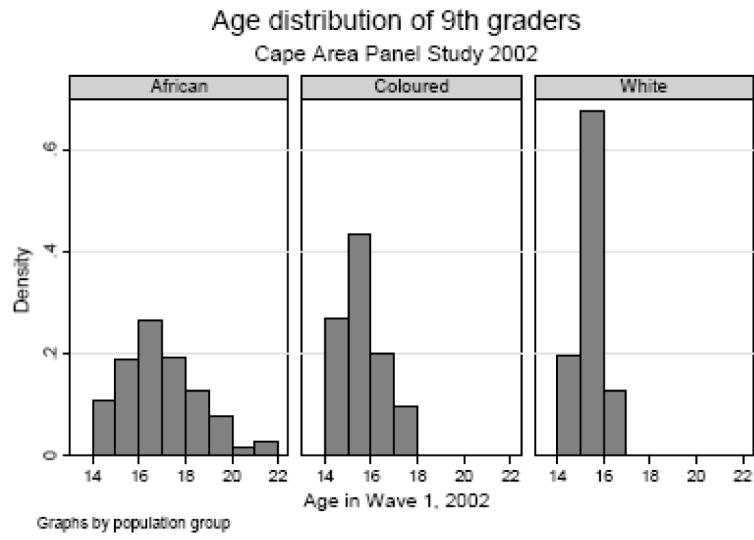
Figure 4.
Age distribution of 9th graders, CAPS Wave 1, 2002
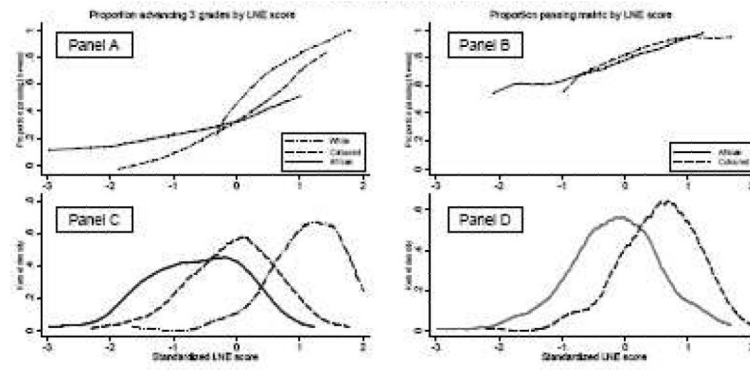
Figure 5.
Lowess estimates of proportion advancing three grades and proportion passing matric by baseline LNE score, with kernel densities of LNE score

Table 1

Sample size by population group and grade and attrition between waves, respondents in grades 8 to 12 in 2002, Cape Area Panel Study

Population Group	CAPS Wave1, 2002			Sample with data for 2005			
	Sample size	Unweighted percent	Weighted percent	Sample size	Unweighted percent	Weighted percent	Rate of attrition
Black/African	1,157	46.7	29.8	1,001	45.8	29.1	13%
Coloured	973	39.3	49.3	919	42.1	52.8	6%
White	349	14.1	20.9	265	12.1	18.1	24%
Total	2,479	100.0	100.0	2,185	100.0	100.0	12%

Note: Sample with data for 2005 was interviewed in either Wave 3 (2005) or Wave 4 (2006)

Table 2

Percentage who worked in last 12 months, CAPS respondents in Wave 1, 2002

Age	African		Coloured		White	
	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

Status in 2005	8th grade in 2002			9th grade in 2002		
	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	29.3	15.5	8.0
Enrolled in Grade 12	2.5	0.7	7.3	27.8	40.6	80.8
Post-secondary	1.6	1.9	0.0	1.3	1.8	3.7
Total enrolled	81.9	63.6	92.5	69.4	61.4	92.4
Not enrolled/not working	15.0	24.0	3.0	22.3	22.9	1.1
Not enrolled/working	3.1	12.5	4.5	8.3	15.7	6.5
Sample size	141	132	41	249	227	58

Table 4

Descriptive statistics, Cape Area Panel Study, respondents enrolled in Grades 8 or 9 in 2002 and observed again in 2005

Variable	African (N=458)		Coloured (N=391)		White (N=84)	
	Mean	Std. Dev.	Mean	Std. Dev.	Mean	Std. Dev.
Advance 3 grades by 2005	0.271	0.445	0.340	0.474	0.816	0.390
Enrolled in 2003	0.898	0.303	0.820	0.385	0.966	0.182
Enrolled in 2004	0.822	0.383	0.685	0.465	0.957	0.204
Enrolled in 2005	0.673	0.470	0.478	0.500	0.843	0.366
Grade 9 in 2002	0.643	0.480	0.632	0.483	0.587	0.495
Female	0.555	0.498	0.461	0.499	0.532	0.502
Number of grades failed by 2002	0.770	0.969	0.553	0.755	0.211	0.492
Failed at least one grade by 2002	0.516	0.500	0.413	0.493	0.175	0.382
Failed grade enrolled in 2002	0.171	0.377	0.162	0.369	0.019	0.139
Standardized LNE total score	−0.626	0.814	−0.028	0.723	1.065	0.602
Log per cap hh income (mean zero)	−0.681	0.891	0.216	0.874	1.871	0.757
Mother's education (grades completed)	8.206	2.954	8.690	2.632	12.284	1.604
Father's education (grades completed)	7.194	3.592	8.762	2.916	12.920	1.994
Mother's education missing	0.107	0.309	0.098	0.297	0.000	0.000
Father's education missing	0.395	0.489	0.318	0.466	0.085	0.281
Local unemployment rate for age & sex	0.773	0.148	0.600	0.206	0.262	0.264
Former DET (African) school	0.748	0.435	0.020	0.139	0.000	0.000
Former HOA (White) school	0.033	0.180	0.074	0.263	0.907	0.293
Former HOR (Coloured) school	0.128	0.335	0.862	0.346	0.012	0.109
New school since 1994	0.091	0.287	0.009	0.092	0.082	0.276
Annual school expenses in 2002 (rands)	325.8	1161.5	731.4	1330.0	5816.5	5848.6
Pupil-teacher ratio in 2002	32.4	4.0	30.4	3.1	24.0	3.1

Table 5

Probit regressions for probability of advancing 3 grades between 2002 and 2005, CAPS respondents in grades 8 or 9 in 2002

Variable	Probit coefficients			Tests for equality of coefficients			Marginal effects at X_1		
	African	Coloured	White	African-Coloured	African-White	White-Coloured	African	Coloured	White
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)
Number of grades failed, 2002	−0.202 [0.083]**	−0.69 [0.165]***	−0.846 [0.501]*	7.089 (0.008)***	1.632 (0.202)	0.086 (0.769)	−0.074 [0.031]**	−0.239 [0.045]***	−0.221 [0.170]
Standardized LNE total score	0.313 [0.087]***	0.743 [0.128]***	0.871 [0.407]**	7.776 (0.005)***	1.826 (0.177)	0.092 (0.760)	0.115 [0.035]***	0.256 [0.057]***	0.227 [0.146]
Log hh income per capita	0.094 [0.086]	0.213 [0.097]**	0.567 [0.295]*	0.850 (0.357)	2.414 (0.121)	1.304 (0.254)	0.035 [0.032]	0.074 [0.035]**	0.148 [0.105]
Mother's schooling	−0.002 [0.032]	0.084 [0.041]**	−0.106 [0.106]	2.776 (0.096)*	0.895 (0.344)	2.845 (0.092)*	−0.001 [0.012]	0.029 [0.014]**	−0.028 [0.035]
Father's schooling	0.033 [0.024]	0.021 [0.032]	0.129 [0.125]	0.090 (0.763)	0.577 (0.447)	0.711 (0.399)	0.012 [0.009]	0.007 [0.011]	0.034 [0.034]
Unemployment rate	0.408 [0.577]	−0.254 [0.391]	−0.035 [0.931]	0.894 (0.345)	0.165 (0.684)	0.048 (0.826)	0.150 [0.205]	−0.088 [0.134]	−0.009 [0.242]
Grade 9 in 2002	−0.150 [0.117]	−0.501 [0.169]***	−0.775 [0.375]**	2.958 (0.086)*	2.569 (0.110)	0.444 (0.505)	−0.053 [0.042]	−0.146 [0.056]***	−0.134 [0.126]
Female	−0.159 [0.151]	0.181 [0.166]	0.14 [0.518]	2.330 (0.127)	0.312 (0.576)	0.005 (0.939)	−0.06 [0.057]	0.059 [0.056]	0.034 [0.121]
Constant	−0.53 [0.536]	−0.713 [0.570]	−0.377 [1.516]	0.054 (0.815)	0.009 (0.923)	0.043 (0.834)			
Predicted probability at X_1							*0.343*	*0.295*	*0.178*
Observations	458	391	84						

Robust standard errors in brackets in columns 1-3 and 7-9; p-value of F tests in parentheses in columns 4-6

*
significant at 10%

NIH-PA Author Manuscript

NIH-PA Author Manuscript

NIH-PA Author Manuscript

**

significant at 5%

***

significant at 1%. Regression includes dummy variables for missing values on parental schooling and household shock. Marginal effects evaluated for female in grade 8 in 2002, 1 failed grade, LNE=0, log income=0, parental schooling=8, unemployment rate=0.6.

Table 6

Probit regressions for probability of enrollment in 2004, CAPS respondents enrolled in grades 8 or 9 in 2002

Variable	Probits for enrollment		Test of equality of coefficients	Marginal effects at X_1	
	African	Coloured		African	Coloured
	(1)	(2)	(3)	(4)	(5)
Number of grades failed, Wave 1	−0.138 [0.077]*	−0.390 [0.106]***	3.773 (0.052)*	−0.031 [0.020]	−0.107 [0.039]***
Standardized LNE total score	0.139 [0.088]	0.258 [0.123]**	0.603 (0.437)	0.031 [0.019]*	0.071 [0.032]**
Log hh income per cap.	−0.086 [0.108]	0.335 [0.107]***	7.667 (0.005)***	−0.020 [0.026]	0.092 [0.036]**
Mother's schooling	0.084 [0.029]***	−0.027 [0.034]	6.468 (0.011)**	0.019 [0.008]**	−0.008 [0.009]
Father's schooling	−0.017 [0.032]	0.089 [0.037]**	4.784 (0.029)**	−0.004 [0.007]	0.024 [0.011]**
Local unemployment rate for age and sex	1.451 [0.549]***	0.170 [0.408]	3.718 (0.054)*	0.328 [0.184]*	0.047 [0.112]
Failed in 2002	−0.696 [0.161]***	−1.420 [0.221]***	7.038 (0.008)***	−0.212 [0.066]***	−0.517 [0.077]***
Grade 9 in 2002	−0.099 [0.168]	−0.346 [0.162]**	1.142 (0.285)	−0.024 [0.039]	−0.109 [0.048]**
Female	0.045 [0.172]	0.035 [0.155]	0.001 (0.965)	0.010 [0.039]	0.010 [0.043]
Constant	−0.248 [0.524]	0.622 [0.499]	1.540 (0.215)		
Predicted probability at X_1				0.857	0.806
Observations	441	394			

Robust standard errors in brackets in columns 1-2 and 4-5; p-value of F tests in parentheses in column 3

NIH-PA Author Manuscript

NIH-PA Author Manuscript

NIH-PA Author Manuscript

*
significant at 10%

**
significant at 5%

***
significant at 1%. Regression includes dummy variables for missing values on parental schooling and household shock. Marginal effects evaluated for female in grade 8 in 2002, 1 failed grade, LNE=0, log income=0, parental schooling=8, unemployment rate=0.6, passed grade in 2002.

Table 7

Pass rates and descriptive statistics, CAPS respondents enrolled in grades 9 to 12 in 2002, 2003 or 2004

Variable	African Mean	African Std. Dev.	Coloured Mean	Coloured Std. Dev.	White Mean	White Std. Dev.
Grade 12:						
Pass grade	0.778	0.417	0.896	0.306	0.996	0.066
Pass with exemption	0.183	0.387	0.229	0.421	0.586	0.494
Number of attempts at grade	1.102	0.329	1.011	0.104	1.000	0.000
Number of grades failed, Wave 1	0.696	0.854	0.404	0.693	0.140	0.362
Standardized LNE total score	−0.089	−0.729	0.564	−0.625	1.311	−0.449
Log hh income per cap.	−0.421	−0.958	0.676	−0.798	1.870	−0.766
Observations	294		350		132	
Grade 11:						
Pass grade	0.644	0.479	0.793	0.406	0.917	0.276
Number of attempts at grade	1.197	0.422	1.070	0.283	1.016	0.127
Number of grades failed, Wave 1	0.782	0.986	0.374	0.669	0.156	0.423
Standardized LNE total score	−0.318	0.782	0.396	0.638	1.228	0.522
Log hh income per cap.	−0.497	0.896	0.604	0.775	1.775	0.937
Observations	477		422		149	
Grade 10:						
Pass grade	0.676	0.468	0.694	0.461	0.932	0.253
Number of attempts at grade	1.212	0.445	1.156	0.378	1.070	0.256
Number of grades failed, Wave 1	0.775	0.913	0.485	0.727	0.157	0.433
Standardized LNE total score	−0.501	0.783	0.151	0.693	1.151	0.554
Log hh income per cap.	−0.617	0.890	0.403	0.846	1.816	0.790
Observations	601		503		125	
Grade 9:						
Pass grade	0.815	0.388	0.826	0.380	0.959	0.199
Number of attempts at grade	1.136	0.374	1.056	0.229	1.015	0.124
Number of grades failed, Wave 1	0.780	0.976	0.532	0.706	0.185	0.466

Variable	African		Coloured		White	
	Mean	Std. Dev.	Mean	Std. Dev.	Mean	Std. Dev.
Standardized LNE total score	−0.642	0.798	−0.036	0.731	1.052	0.589
Log hh income per cap.	−0.690	0.897	0.248	0.869	1.845	0.700
Observations	531		394		96	

Table 8

Probit regressions for probability of passing Grade 12 matriculation exam and probability of passing Grades 9, 10, and 11

Variable	Probit coefficient		Test of equality of coefficients	Marginal effects at X₁	
	African	Coloured		African	Coloured
	(1)	(2)	(3)	(4)	(5)
Grade 12:					
Number of grades failed, 2002	−0.059 [0.106]	−0.326 [0.120]***	1.931 (0.164)	−0.017 [0.030]	−0.080 [0.036]**
Standardized LNE total score	0.442 [0.114]***	0.486 [0.172]***	0.038 (0.843)	0.127 [0.039]***	0.120 [0.047]**
Log hh income per capita	0.293 [0.098]***	0.274 [0.125]**	0.015 (0.901)	0.084 [0.032]***	0.067 [0.034]*
Predicted probability at X₁				0.792	0.837
Observations	327	352			
Grade 11:					
Number of grades failed, 2002	−0.207 [0.050]***	−0.256 [0.103]**	0.188 (0.664)	−0.079 [0.020]***	−0.081 [0.037]**
Standardized LNE total score	0.189 [0.078]**	0.34 [0.129]***	0.626 (0.428)	0.073 [0.029]**	0.108 [0.045]**
Log hh income per capita	0.037 [0.075]	0.177 [0.109]	0.881 (0.347)	0.014 [0.029]	0.056 [0.037]
Predicted probability at X₁				0.611	0.751
Observations	566	452			
Grade 10:					
Number of grades failed, 2002	−0.038 [0.059]	−0.203 [0.093]**	1.545 (0.213)	−0.013 [0.019]	−0.077 [0.037]**

Variable	Probit coefficient		Test of equality of coefficients	Marginal effects at X_1	
	African	Coloured		African	Coloured
	(1)	(2)	(3)	(4)	(5)
Standardized LNE total score	0.179 [0.069]***	0.43 [0.093]***	2.966 (0.084)*	0.058 [0.021]***	0.163 [0.033]***
Log hh income per capita	−0.034 [0.060]	0.302 [0.089]***	7.512 (0.006)***	−0.011 [0.019]	0.114 [0.035]***
Predicted probability at X_1				*0.737*	*0.626*
Observations	729	581			
Grade 9:					
Number of grades failed, 2002	−0.087 [0.067]	−0.404 [0.121]***	4.150 (0.041)**	−0.016 [0.014]	−0.100 [0.037]***
Standardized LNE total score	0.276 [0.092]***	0.312 [0.131]**	0.038 (0.844)	0.052 [0.018]***	0.077 [0.035]**
Log hh income per capita	0.163 [0.085]*	0.033 [0.111]	0.669 (0.413)	0.031 [0.014]**	0.008 [0.028]
Predicted probability at X_1				*0.890*	*0.835*
Observations	602	415			

Robust standard errors in brackets in columns 1, 2, 4, and 5; p-value of F tests in parentheses in column 3

*
significant at 10%

**
significant at 5%

***
significant at 1%. Sample for each regression is the pooled sample of respondents enrolled in a given grade in 2002, 2003, or 2004; regressions also include variables used in Table 5; Marginal effects evaluated for female, 1 failed grade, LNE=0, log income=0, parental schooling=8, no shock, unemployment rate=0.6.