Explaining the persistence of racial gaps in schooling in South Africa

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

Nicola Branson
University of Cape Town
Nicola.Branson@uct.ac.za

David Lam
University of Michigan
davidl@umich.edu

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

Abstract

This paper analyses the large racial differences in progress through secondary school in South Africa using recently collected longitudinal data. Following the progress of students who were enrolled in Grades 8 and 9 in 2002 in the Cape Area Panel Study, we document large differences in the probability of grade advancement between white, coloured, and African youth. Probit regressions indicate that grade advancement between 2002 and 2005 is strongly associated with household income and with respondents' scores on a baseline literacy and numeracy test. We fully explain the white and coloured advantage over Africans in

progress through school when we control for baseline test scores, previous grades failed, and per capita household income. The results suggest that the early disadvantage of African secondary students is a major factor driving poor progress through secondary school, with continued racial gaps in grade progression contributing to persistent racial gaps in ultimate schooling attainment. These key results do not change when we re-estimate these equations separately by race and conduct our statistical tests across these equations or even when we conduct post-estimation counterfactual simulations or propensity score matching. As a final check we add a set of school specific factors to the probit regressions by race. These factors are not statistically significant in the models; their introduction only marginally reduces the importance of the baseline test scores and previous grades failed and they are not important in the counterfactual analysis. All in all the paper provides very strong evidence that knowledge accumulated by Grade 8 or 9 is a critical determinant of progress through secondary school and that the equalizing of secondary school quality is unlikely to eliminate racial gaps in grade progression without improvements at earlier grades.

1 Introduction

More than a decade after the end of apartheid, South Africa continues to experience large racial differences in schooling attainment. While there has been a great deal of progress in access to education and equalization of government expenditures on schools, there continue to be large disparities in the quality of schools attended by students from different racial groups (Fiske and Ladd, 2004, Yamauchi, 2005, Van der Berg, 2007). Primary schooling is almost universal for all racial groups and enrolment rates are high into the teenage years, yet there are large racial differences in the rate at which students progress through school and in the proportion of students that complete secondary school (Anderson *et al.* 2001, Lam *et al.* 2011).

This paper looks at a critical period in the schooling experience of young people – the years following Grades 8 and 9. In the South African School system, Grade 8 is the first grade of secondary school. We use the Cape Area Panel Study (CAPS), a new panel study of youth collected in Cape Town, to follow 8th and 9th Graders for the next three years. We find large racial differences in the probability that students successfully advance three grades in school between 2002 and 2005. While 82% of white students advance three grades during this three-year period, only 34% of coloured students and 27% of African students advance three grades.

In this paper, we set out to advance our understanding of this differential performance by race. By combining our CAPS data with a school-level data set as well as a set of community variables from the 2001 census, we are able to look at the influence of a large set of individual, household, school, and community variables in our analysis. These variables include previous school outcomes, performance on a baseline literacy and numeracy evaluation, household level variables such as income and parental schooling, measures of school quality such as student-teacher ratios, and community measures such as neighbourhood unemployment rates. Given that the existing

South African literature has been restricted to examining the determinants of educational performance from the perspective of school-level data, the results from our data set are potentially valuable in that they give a new perspective on this issue.

After describing the CAPS data in Section 2, the rest of the paper provides four empirical views on the relative importance of these variables on progress through school. Section 3 begins by presenting some descriptive results that are important in highlighting the starkness of racial differences in most variables. This is important in highlighting differences by race. It is also important in making clear the extent of racial mixing in schools and the potential role of school choice on our empirical work. In Section 4 we pool all racial groups and estimate a series of probit regressions. We start by regressing grade advancement on a set of racial dummy variables. We then add sets of variables to this basic regression model in order to assess how each set changes the magnitude of the race coefficients. These changes are very interesting. However, as shown in descriptive statistics, there is very little overlapping of the distributions by race for many of the covariates that are important in explaining progress through school. This makes the support for pooling tenuous. Therefore, in Section 5 we estimate a series of probit equations separately by race and compare results across equations. We then draw on some recent work by Cameron and Heckman (2001) to conduct post-estimation counterfactual simulations that take African or coloured or white learners with their characteristics and trans-plant them into the equations of other racial groups.

The legacy of apartheid lingers in Cape Town's schools in the sense that they remain strongly segregated by race. There are two major reasons for this: First,the city remains strongly geographically segregated by race, with large distances often separating learners of one race from schools attended by learners from other races. Second, post-apartheid education policy has given schools considerable latitude to set school fees, with CAPS learners exhibiting considerable variance in the school fees charged by schools. The result is that learners often attend the same schools that were built in their neighbourhoods under the race-based apartheid education system. In Section 6 we add a number of variables capturing these school types and school fees in order to see how they impact progress through school.

Our final approach in Section 7 uses propensity score matching. By secondary school the amount of accumulated disadvantage, including factors which directly affect progress through secondary school, differs by race. Thus, in order to isolate the effect of race on progress through secondary schooling, we compare students with similar 2002 baseline characteristics from different racial groups.

Our empirical work shows that a small set of variables, including baseline test scores, household income, and parental education, can account for the large racial difference in grade advancement. The results suggest that the human capital students bring with them into high school largely determines their chances of completing high school.

While it is important that African, coloured, and white students experience vastly different levels of resources in the high schools they attend, we find that these differences are less important in explaining high school success than the skills the students bring into high school. Put another way, while there is appropriate concern about the large quality differences in high schools, our results suggest that even if African and white students were to attend identical high schools, there would still be large racial differences in grade advancement. Our matching analysis shows that equalizing characteristics between racial groups at the start of secondary school would eliminate the racial difference in progress through secondary school between coloureds and whites and result in an African advantage in the gap when compared to coloured students.

2 The Cape Area Panel Study

Our analysis is based on the Cape Area Panel Study (CAPS), a longitudinal study of youth and their families in metropolitan Cape Town. Details about the design of CAPS, a collaborative project of the University of Cape Town and the University of Michigan, are available in Lam *et al* (2008).[1] Wave 1 of CAPS, which was collected in 2002, included 4,752 young people aged 14-22, living in 3,304 households. CAPS was designed as a stratified two-stage clustered sample with stratification on the predominant population group living in each sample cluster. Cape Town has three predominant population groups – coloured, African/Black, and white. The

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

Wave 1 of CAPS contains two major sources of data. First, the survey includes a household questionnaire, in which demographic data on the entire household is collected. Second, the survey includes a detailed young adult questionnaire, which collects data on schooling, employment, and fertility of household members between the ages of 14 and 22. It also includes basic numeracy and literacy skills tests administered to each youth respondent. The results of this test will be used in the analysis below. CAPS youth respondents were interviewed a second time in either 2003 or 2004, and were interviewed a third and fourth time in 2005 and 2006 respectively. Parental schooling information is collected from the household questionnaire when the parent is co-resident, and from the young adult directly when the parent is not co-resident.

The Wave 1 (2002), Wave 3 (2005) and Wave 4 (2006) provide the data for

---

1. Technical documentation and background information is available on the CAPS web site, *www.caps.uct.ac.za* .

the analysis in this paper. The base wave of CAPS collected information on years of completed schooling by 2002 and actual school enrolments in 2002, Thereafter, each time a young adult was re-interviewed the record of progress through school was updated as from the last time that the young adult was interviewed. Respondents who were enrolled in Grades 8 or 9 in 2002 are the major focus of this paper. Information about their progress through school between 2002 and 2005 could have been gathered in Wave 3 or Wave 4. For this reason we will speak of Wave 3/4 in the rest of the paper.

Table 1 shows the sample size in Wave 1 and Waves 3/4 for respondents who were enrolled in Grades 8 and 9 in 2002. As seen, there were just over 1,000 respondents in Grades 8 and 9 in Wave 1, 48% of whom were African. The "weighted percent" column shows that, when we adjust for the oversampling of African respondents, the African group is 32% of those enrolled in Grades 8 and 9. The white sample is considerably smaller, a result of both the intentional sample design and the lower response rate among white households. CAPS response rates were high in African and coloured areas and low in white areas. This is in line with most South African household surveys, Household response rates were 89% in African areas, 83% in coloured areas, and 46% in white areas. Young adult response rates, conditional on participation of the household, were quite high, even in white areas. Given household participation, response rates for young adults were 93% in African areas, 88% in coloured areas, and 86% in white areas (Lam *et al*, 2008).

As shown in Table 1, the overall rate of attrition between Wave 1 and

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

Population Group	CAPS Wave1, 2002			CAPS Wave 3, 2005 or CAPS Wave 4, 2006			Rate of attrition
	Sample size	Unweighted percent	Weighted percent	Sample size	Unweighted percent	Weighted percent	
Black/African	515	48.27	31.69	461	47.62	31.25	10%
Coloured	423	39.64	50.89	401	41.43	52.72	5%
White	129	12.09	17.43	106	10.95	16.03	18%
Total	1067	100.0	100.0	968	100.0	100.0	9%

Wave 3 was 9%, with significant differences across population groups. The African attrition rate is 10%, with most of the attrition resulting from migration back to the rural Eastern Cape province that is the main sending region for Africans living in Cape Town. The coloured population has its roots primarily in Cape Town, a factor contributing to its lower 5% attrition rate. The 18% attrition rate for whites includes both migration out of Cape Town (including migration out of South Africa) and a significant number of refusals.

The major focus of this paper is the comparison of schooling outcomes for African, coloured, and white youths. These three population groups were subject to very different treatment under apartheid. Many of these apartheid-era differences are likely to continue affecting young people in the post-apartheid period. Whites had advantages in a wide range of areas, including significantly higher expenditures on schooling, privileged access to the labour market, unrestricted residential mobility, and better access to most social services. Africans had the least access to services and the most restrictions on work and migration, with a large gap in expenditures on schooling. The coloured population, which is heavily concentrated in Cape Town, occupied an intermediate status under apartheid, with higher expenditures on schooling, fewer restrictions on residential mobility, and better access to jobs. As mentioned earlier, residential de-segregation has been very slow and new schools are no less racially distinct than schools existing prior to 1994. For example, all new schools attended by African respondents are located in African townships. Our CAPS data gives a fuller descriptive picture of the situation faced by Cape Town's youth after 2000 and we now turn to these data.

3 Initial empirical evidence

Using the CAPS data from Waves 1, 2, 3 and 4, we can follow the progress through school of young people who were enrolled in school in 2002. Students who were in Grades 8 and 9 in 2002 and who remained in school for the next three years and passed all grades would have reached Grades 11 and 12, respectively, by 2005. Table 2 shows the activities in 2005 of those who were in Grade 8 and Grade 9 in 2002 at the time of the survey, disaggregated by population group. About 82% of whites in this group advanced three grades in three years. The experience of African and coloured youth is very different. Among Africans only 27% advanced three grades in three years and the equivalent percentage for coloured is 34%. Clearly, those that drop out of school do not pass three grades in three years. Thus, dropout rates are important too. The enrolment figures in Table 2 make it clear that white pupils rarely drop out and that African youth are much more likely to stay in school than coloured youth; this in spite of the fact that they have higher rates of grade repetition.

Table 2 Descriptive statistics, Cape Area Panel Study Waves 1-4, CAPS respondents enrolled in Grades 8 or 9 in 2002 and observed again in 2005 or 2006

Variable	African (N=452)		Coloured (N=385)		White (N=83)	
	Mean	Std. Dev.	Mean	Std. Dev.	Mean	Std. Dev.
Advance 3 grades by 2005	0.269	0.444	0.339	0.474	0.822	0.385
Enrolled in 2003	0.897	0.304	0.817	0.387	0.966	0.182
Enrolled in 2004	0.824	0.381	0.690	0.463	0.957	0.205
Enrolled in 2005	0.678	0.468	0.485	0.500	0.848	0.361
Grade 9 in 2002	0.641	0.480	0.629	0.484	0.584	0.496
Female	0.559	0.497	0.463	0.499	0.530	0.502
Age	16.2	1.7	15.0	1.1	14.6	0.7
Number of grades failed by 2002	0.766	0.969	0.549	0.755	0.212	0.494
Failed at least one grade by 2002	0.513	0.500	0.409	0.492	0.176	0.383
Failed grade enrolled in 2002	0.167	0.373	0.164	0.371	0.019	0.139
Standardized LNE total score	-0.624	0.815	-0.033	0.723	1.072	0.597
Log per cap hh income (mean zero)	-0.679	0.893	0.212	0.876	1.871	0.759
Mother's education (grades completed)	8.2	3.0	8.7	2.6	12.3	1.6
Father's education (grades completed)	7.2	3.6	8.8	2.9	12.9	2.0
Mother's education missing	0.108	0.310	0.097	0.296	0.000	0.000
Father's education missing	0.394	0.489	0.318	0.466	0.079	0.272
Local unemployment rate for age & sex	0.774	0.149	0.602	0.205	0.263	0.264
Former DET (African) school	0.747	0.435	0.020	0.140	0.000	0.000
Former HOA (White) school	0.031	0.173	0.072	0.260	0.919	0.274
Former HOR (Coloured) school	0.131	0.338	0.865	0.342	0.010	0.101

New school since 1994	0.091	0.288	0.009	0.093	0.070	0.258
Annual school expenses in 2002 (rands)	324.05	1166.77	730.48	1334.57	5839.95	5860.59
Pupil-teacher ratio in 2002	32.4	4.0	30.4	3.1	24.0	3.0
Moved school between 2002 and 2003	0.203	0.403	0.131	0.338	0.024	0.153
Moved school between 2003 and 2004	0.117	0.321	0.034	0.181	0.084	0.280
Moved school between 2004 and 2005	0.191	0.394	0.136	0.344	0.204	0.406
Moved former department 2002-2003	0.048	0.214	0.017	0.130	0.000	0.000
Moved former department 2003-2004	0.042	0.201	0.003	0.058	0.000	0.000
Moved former department 2004-2005	0.024	0.154	0.024	0.152	0.104	0.308

In the regressions below we use a number of individual, household, and community characteristics to predict progress through school. Table 2 goes on to provide an overview of some of these characteristics. One interesting feature of CAPS is the numeracy and literacy evaluation that was administered to all youth respondents in Wave 1. This was a self-administered written test that was taken by respondents after the completion of the young adult questionnaire. The test had 45 questions and took about 20 minutes to complete. The respondent could choose to take the test in either English or Afrikaans. There was no version in Xhosa, the home language of most African respondents. The English language test was taken by 99% of the African respondents, 43% of the coloured respondents, and 64% of the white respondents. In interpreting the results below it is important to keep in mind that most African respondents took the test in a second language, while white and coloured students took the test in their first language. We use the test below as a measure of cumulative learning at the time of the interview. Performance on the test reflects a combination of many factors, including innate ability, home environment, and the quantity and quality of schooling up to that point.

The left hand panel of Figure 1 presents kernel density estimates of the distribution of combined (total) scores on the numeracy and literacy tests for each population group. Each score is standardized to zero mean and unit variance. The differences in test scores across population groups are striking.

We see only a small area of overlap between the test scores of African and white respondents. The distribution of test scores for coloured youth sits between, with considerable overlap with both the white and African distributions. The mean standardized test score is -0.62 for Africans, -0.03 for coloureds, and 1.07 for whites, implying a standard deviation gap between whites and Africans that is close to 2. The right hand panel of Figure 1 shows the proportion of students progressing three grades in three years against their score. It is clear from the graph that the gradient of the line is different across race. Africans, followed by coloured have the flattest curve, indicating that the test score is less predictive of progress through school. The curves intersect at the average score, signifying no differences between races in progression through secondary school. Africans are more likely to progress given scores below the mean, and less likely to progress given scores above the mean. This large difference in the distribution of test scores is important to keep in mind in our regressions below, where we will include the test scores as regressors.
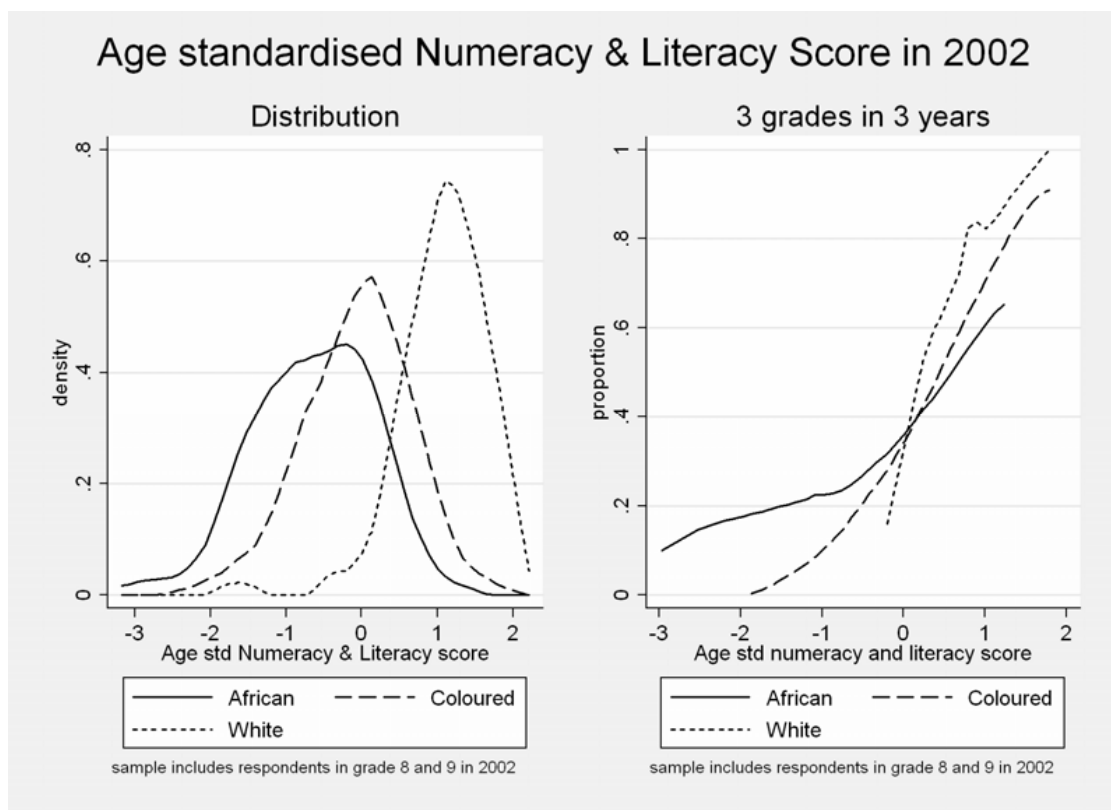

Figure 1

We are also very interested in the impact of household income on grade progression, and will focus in our simulations on the extent to which racial differences in income can explain racial differences in grade progression. We

use the log of per capita household income in 2002, the year in which we first observe the students, as reported by an adult respondent in the Wave 1 household questionnaire. Figure 2 plots the kernel densities for the distribution of income for each population group, standardized to the mean income for the combined population in the left hand panel and the probability of advancing three grades in three years in the right hand panel. Once again we see very large differences between population groups. The difference in mean log income between whites and Africans is about 2.6. Exponentiated, this implies that white youth in 2002 were living in homes with over 10 times higher per capita household income than Africans. As was the case with test scores, a striking feature of Figure 2 is the very small range in which the African and white income distributions overlap. The coloured distribution sits between the two distributions, overlapping more with the African distribution than with the white distribution.
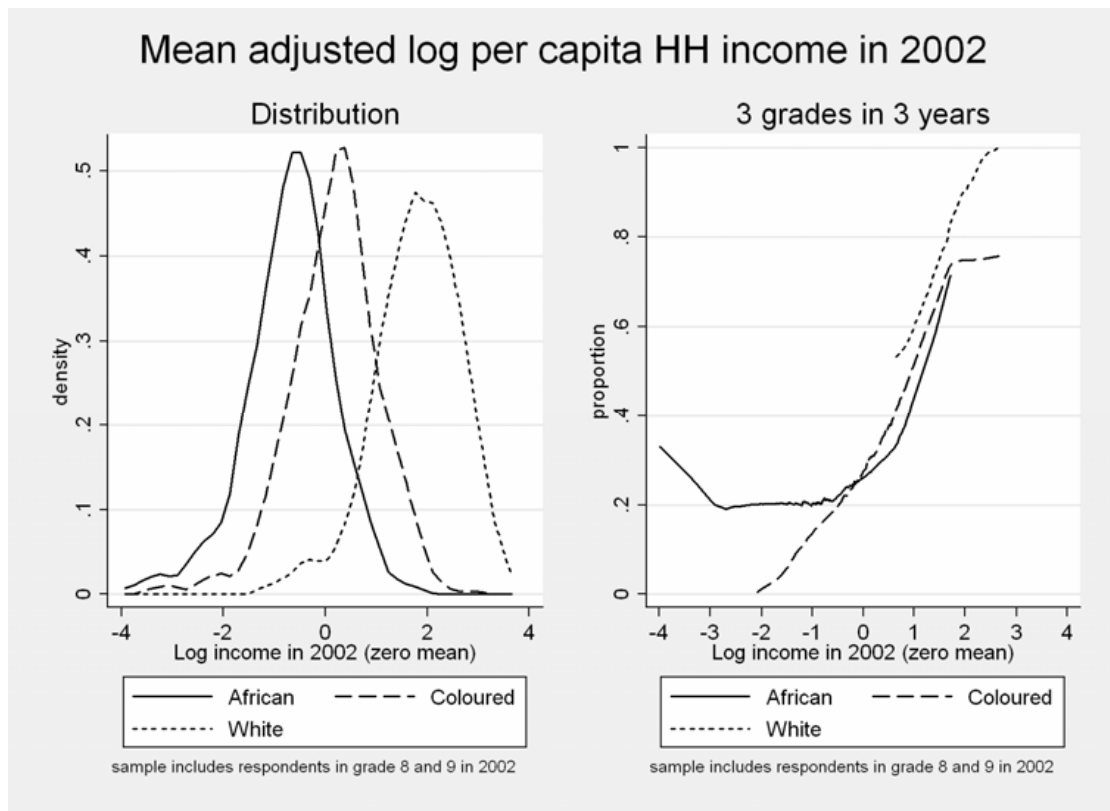
Figure 2

Another important variable to consider in explaining progress through school is the extent to which students were already behind in school in 2002. 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 3 shows the age distribution for 8th and 9th Graders in 2002. Looking at 8th Graders in the top panel, we see that there is probably

some truncation due to the fact that our sample begins at age 14. We lose some 13 year-olds who would have been in Grade 8 in 2002, though this will have been a small proportion of all 8th Graders. There are large differences in the age distribution of 8th Graders across population groups. White 8th Graders are concentrated at age 14, with less than 20% at age 15.

By contrast, the modal age of African 8th Graders is 15, with a wide distribution ranging between ages 14 and 21. The differences are even greater among 9th Graders. About 90% of white 9th Graders are age 15 or below, compared to 29% of African 9th Graders and 70% of coloured 9th Graders. Roughly 25% of African 9th Graders are age 18 or older.

Age distribution of 9th graders
Cape Area Panel Study 2002

Age in Wave 1, 2002
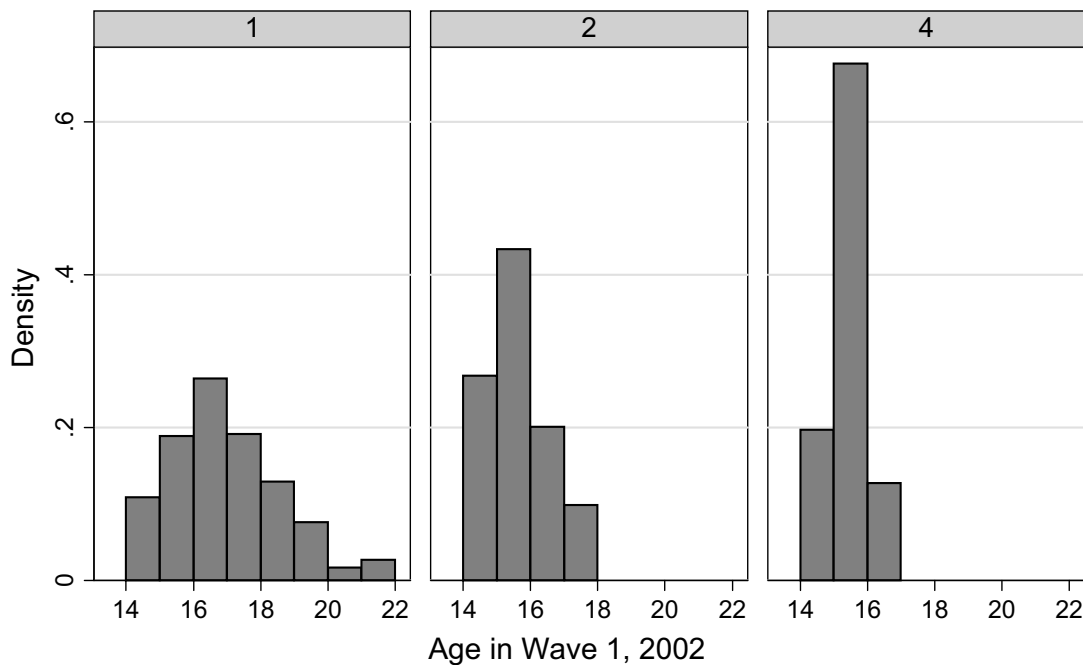
Graphs by population group

Figure 3

Table 2 also presents useful background information about school characteristics, school fees and school mobility. The variables for former department indicate that 75% of African youth attend schools that were classified as African schools (Department of Education and Training) under apartheid. About 13% attend formerly coloured (House of Representatives) schools,

3% attend formerly white (House of Assembly) schools, and 9% attend schools that were created since 1994 and hence have no "former department" classification. Note that 87% of coloured students are in formerly coloured schools and 92% of white students are in formerly white schools. The annual school expenditure variable shows the enormous differences in

school fees. African students paid an average of 324 rands per year (roughly 32 dollars), coloured students paid 730 rands, and white students paid 5,840 rands. Since these fees are often used to hire extra teachers, the differences in fees translate into differences in pupil-teacher ratios. This is seen in the next row of Table 2, which shows a mean pupil-teacher ratio of 32.4 for Africans, compared to 24 for whites. There is some movement between schools, with 19% of Africans changing schools between 2004 and 2005. For all races respondents are most likely to change schools after Grade 8. Although a fair number of respondents change schools, very few move to a school that fell under a different department before 1994. Between 2004 and 2005, when most school changes occurred, only 2.4% of Africans, 2.4% of coloureds and 10.4% of whites changed former department.

Finally, we merged a set of local level statistics (community variables) from the census into the CAPS data. The one variable from this merge that we use in this paper are local unemployment rates for those of the same age and gender. It brings into the model some sense of the opportunity cost of being in school and focusing on progress through school. However, this is a community-level variable and communities with high unemployment rates are not conducive to studying. Thus, *a priori*, there is some ambiguity about the impact of this variable on progress through school.

4 Pooled probit regressions of who passes three grades in three years

This section presents results of probit regressions in which our dependent variable is an indicator of progress through school between 2002 and 2005 for our sample of 8th and 9th Graders. Our dependent variable is equal to 1 if the respondent advanced at least three years in school by 2005 – 8th Graders reached at least Grade 11 and 9th Graders reached at least Grade 12. The dependent variable is equal to 0 if there is any other outcome, including dropping out of school before reaching the target grade or being in school in some grade below the target grade in 2005. Table 3 presents the probit regressions for grade advancement of 8th and 9th Graders. For each regression we report the marginal effect evaluated at the sample means with the robust standard error presented below it in brackets. Regression (1) includes a dummy for African and a dummy for coloured, with white as the omitted category. It also controls for the gender of the students by including a female dummy variable. It then includes a dummy variable for Grade 9, which allows the probability of advancement from Grade 9 to Grade 12 to have a different mean than the probability of advancement from Grade 8 to Grade 11.

The probit in regression (1) serves as the base case against which the other regressions in Table 3 are compared. The coefficients on African and coloured are both strongly negative and highly significant implying that Africans and coloureds had a very much lower probability of advancing three grades than white respondents between 2002 and 2005. In the rest of the regressions in Table 3 we will see how these popu-

lation group differences are affected by controls for individual, household, and school characteristics.

Table 3 Probit regressions (marginal effects) for probability of advancing 3 grades between 2002 and 2005, CAPS respondents in Grades 8 or 9 in 2002

	(1)	(2)	(3)	(4)
ref: White				
African	-0.49***	-0.14*	-0.1	0.14
	[0.05]	[0.08]	[0.09]	[0.10]
Coloured	-0.48***	-0.25***	-0.17**	-0.05
	[0.06]	[0.08]	[0.08]	[0.08]
ref: Males in Grade 8 in 2002				
Grade 9 in 2002	-0.11***	-0.14***	-0.12***	-0.13***
	[0.04]	[0.04]	[0.04]	[0.04]
Female	0.01	-0.01	0.04	0.02
	[0.03]	[0.04]	[0.04]	[0.04]
Schooling performance up until 2002:				
Number of grades failed, Wave 1		-0.18***		-0.18***
		[0.03]		[0.03]
Standardized LNE total score		0.24***		0.20***
		[0.03]		[0.03]
Household characteristics in 2002:				
Log hh income per cap.			0.12***	0.07***
			[0.02]	[0.02]
Mother's schooling			0.02**	0.01
			[0.01]	[0.01]
Mother's schooling missing			0.22**	0.14
			[0.11]	[0.11]
Father's schooling			0.02***	0.01**
			[0.01]	[0.01]
Father's schooling missing			0.07	0
			[0.08]	[0.07]
Unemployment rate			0.03	-0.06
			[0.11]	[0.11]
Observations	920	920	920	920

Robust standard errors in brackets

* significant at 10%; ** significant at 5%; *** significant at 1%

Regression (2) adds controls for two indicators of the respondent's cumulative learning in 2002; the standardized total score of students on the literacy

and numeracy tests (LNE) and the number of grades failed by 2002. We include the number of grades failed as an indicator of previous poor performance in school, which may also be a reflection of low school quality. The number of grades failed variable has a negative and statistically significant effect on the probability of advancing three years. The coefficient on the standardized LNE score is positive and statistically significant at the 1% level. The magnitude of this coefficient is large, with a one standard deviation increase in the total score increasing the probability of advancing three grades by 24 percentage points.

The change in the African and coloured coefficients from (1) to (2) are striking. The marginal effect point estimate on the African dummy variable drops from -0.49 to -0.14 and is now only significant at the 10% level. The marginal effect on the coloured variable declines from -0.48 to -0.25 and remains statistically significant at the 1% level. Both of these point estimates remain negative compared to the white default. However, the coloured coefficient is now more strongly negative than the African coefficient and the difference between these coefficients is statistically significant. Taken literally, the coefficient implies that if we hold constant test scores and the number of grades behind in school, coloured 8th and 9th Graders were *less* likely to advance three grades in school between 2002 and 2005 relative to both white and African learners.

Regression (3) adds to the base case of (1) a set of variables that is related to household resources and family background as well as a local unemployment

rate. Father's schooling and mother's schooling are measured in years of completed schooling. In cases in which parental schooling is missing, the mother's schooling and father's schooling variables are coded as zero and the parental schooling missing variable is set to one. Mother's schooling is missing for 9% of the observations in Table 2; father's schooling is missing for 33% of the observations. Log of per capita household income is the same variable shown in Figure 2 and discussed above. As mentioned in the previous section of the paper, the local unemployment rate is derived from census data and reflects the unemployment rate for individuals without matric residing in the same small statistical area as the young adult.

Looking at the coefficients in (3), per capita household income has a statistically significant positive effect on the probability of grade advancement. Mother's schooling and father's schooling have small positive effects that are statistically significant at the 5% and 1% level respectively. The local unemployment rate has a small negative coefficient but it is not statistically significant.

Again, as in (2), the impact of the inclusion of these household and community variables on the race dummies is notable. The coefficient on both the African and coloured dummies remain negative but they are much smaller than in (1) and only the coloured coefficient remains statistically significant at the 5% level. Taken literally, they imply that if African students had the same test scores, parental schooling, household income, and other regression (3) characteristics as white students, there would be no discernable difference

between these students in their progress through school between 2002 and 2005.

Regression (4) combines the base case of (1) with the student-level schooling background variables of (2) and the household and community levels variables of (3) in a single estimation. Most of the variables that were included in (2) and (3) variables have very similar coefficients in this combined estimation. Mother's schooling is no longer statistically significant. Again it is the race coefficients that are the most notable. The African coefficient is now positive although insignificant and the coloured coefficient is very small and not statistically significant. Speaking to the African coefficient, at face value this implies that African students with the same test scores, parental schooling, household income, and other regression (2) and (3) characteristics as white students would have a *higher* probability of progressing three grades in three years between 2002 and 2005.

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

do not do better than African students at moving through the higher grades. Potentially this has important policy implications in suggesting that the large differences in secondary school grade progression between population groups cannot be eliminated simply by policies focused on these grades. The cumulative disadvantages that students bring with them into Grades 8 and 9 have powerful impacts on their ability to advance through to the final grades.

Given the weaknesses of the pooled approach, it is fortunate that contemporary empirical approaches allow us to make further progress in teasing out what underlies racial differences in school advancement. With the results from these probits as the backdrop, the next section turns to one of these recent approaches.

5 Probit regressions by race of who passes three grades in three years

The preceding section explored the association between race and progress through school and the extent to which racial differences in progress could be explained by a small number of individual, household and school level characteristics. This section further investigates the role of racial differences in characteristics and extends the analysis to consider racial differences in behaviours. Following Cameron and Heckman (2001), we begin by estimating separate regressions for each of our three population groups and testing for equality of coefficients for each pairwise combination of races. We then use the coefficients to estimate the extent to which differences in characteristics (Xs) and differences in behaviours (βs) can explain the large racial differences in progress through school.

Table 4 Probit regressions (marginal effects) for probability of advancing 3 grades between 2002 and 2005, CAPS respondents in grades 8 or 9 in 2002

	Marginal Effects from Probit Regressions			Tests for Equality of Coefficients		
	African	Coloured	White	African - Coloured	African - White	White - Coloured
	(1)	(2)	(3)	(4)	(5)	(6)
Grade 9 in 2002	-0.07*	-0.15**	-0.1	1.39	1.55	0.38
	[0.04]	[0.06]	[0.07]	(0.24)	(0.21)	(0.54)
Female	-0.05	0.07	-0.01	2.58	0.04	0.25
	[0.05]	[0.06]	[0.08]	(0.11)	(0.84)	(0.62)
Schooling performance up until 2002:						
Number of grades failed, Wave I	-0.07***	-0.23***	-0.15*	5.9	1.92	0.22
	[0.03]	[0.05]	[0.09]	(0.02)	(0.17)	(0.64)
Standardized LNE total score	0.11***	0.24***	0.14**	6.64	1.6	0.11
	[0.03]	[0.04]	[0.06]	(0.01)	(0.21)	(0.74)
Household characteristics in 2002:						
Log hh income per cap.	0.03	0.07**	0.07	1.11	1.48	0.53
	[0.03]	[0.03]	[0.05]	(0.29)	(0.22)	(0.47)
Mother's schooling	0	0.02*	-0.01	2.16	0.72	2.33
	[0.01]	[0.01]	[0.02]	(0.14)	(0.4)	(0.13)
Mother's schooling missing	-0.03	0.31*		2.63		
	[0.11]	[0.18]		(0.11)		

	(1)	(2)	(3)	(4)	(5)	(6)
Father's schooling	0.01 [0.01]	0.01 [0.01]	0.02 [0.02]	0 (0.97)	0.6 (0.44)	0.56 (0.46)
Father's schooling missing	0.02 [0.07]	-0.04 [0.11]	0.07 [0.10]	0.22 (0.64)	0.16 (0.69)	0.26 (0.61)
Unemployment rate	0.1 [0.18]	-0.08 [0.13]	0.01 [0.15]	0.66 (0.42)	0.05 (0.82)	0.11 (0.74)
Constant				0.23 (0.63)	0.03 (0.86)	0.18 (0.67)
Observations	452	385	83			

Robust standard errors in brackets in columns 1-3, p-value of F tests in parentheses in columns 4-6

* significant at 10%; ** significant at 5%; *** significant at 1%

The pooled regressions in Table 3 assume that the responses to the same variables are homogenous across population groups. This assumption is explored in Table 4 where separate regressions for each population group together with tests for the equality of coefficients are presented. The specification corresponds to Regression (4) from Table 3. Columns 1 to 3 of Table 4 present marginal effects estimated at the sample means for each separate sample. Previous grades failed has a substantially less negative effect on grade advancement for African students than for coloured and white students. At the sample means, having failed one additional grade by 2002 is associated with an 8 percentage point lower probability of advancing three grades for Africans, compared to a 22 percentage point lower probability for coloureds and a 15 percentage point lower probability for whites. The LNE score also has a smaller positive effect for African students. At the sample means a one standard deviation increase in the LNE score is associated with an 11 percentage point increased probability of advancing three grades for Africans, compare to a 24 percentage point increase for coloureds. The impact of log per capita household income is not statistically significantly for Africans, but is strongly positive for coloured and white students. There is a significant positive effect of mother's education for coloured students only.

In Columns 4-6 of Table 4 we test for the equality of coefficients between pairs of racial groups across regressions. As shown in Column 4, we can reject the hypothesis that Africans and coloureds have equal coefficients on

previous grades failed, the LNE test, and the mother's schooling missing indicator. This indicates that differences in response to the same variables play an important part in explaining differences in progress through school. 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.

In a previous paper (Lam, Ardington, and Leibbrandt 2011), we provide a theoretical explanation for why the effects on grade progression of characteristics such as the LNE score and household income are smaller for Africans than for coloureds and whites. We argue that African schools do a poor job of evaluating actual learning, resulting in a large stochastic component to grade advancement for Africans. This larger stochastic component weakens the relationship between characteristics and grade progression for Africans compared to coloureds and whites. For purposes of this paper we simply note that these coefficients differ by race, and do our counterfactual analysis using both sets of coefficients for any race pair.

Looking at other variables in our probit in Table 4, we find no significant differences in grade advancement of males and females. This is consistent with 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 only mother's schooling for coloureds significant at the 10% level. This is surprising

given the high variance in parental schooling in our sample and the wide range of research that finds strong effects of parental schooling on children's schooling outcomes. For Africans the coefficient on father's education becomes significant at the 10% level when the LNE scores and number of grades failed are omitted. The coefficient on mother's schooling continues to be insignificant. For coloured students we estimate highly significant positive effects of both parent's schooling when the previous performance outcomes are omitted.

The neighbourhood unemployment rate is not significant for any racial group. It is included here as an attempt to capture two possible effects. On the one hand, the opportunity cost of time may affect either effort in school or the probability of dropping out. On the other hand, better employment prospects might stimulate young people to stay in school and work harder in school. These effects may be cancelling out in our data, although it is also possible that our use of census subplaces does not capture the appropriate labour market. While white and coloured youth appear to have much better job opportunities than African youth due to geographical proximity, family networks, and language skills, there may not be sufficient geographical variation in job opportunities within racial groups to identify an effect.

http://aps.journals.ac.za

Table 5 Predicted change in the racial gap in grade advancement from changing coefficients across samples and from changing values of individual coefficients.

	African equated to coloured	Coloured equated to African	African equated to white	White equated to African	Coloured equated to white	White equated to coloured
	(African β's)	(coloured β's)	(African β's)	(white β's)	(coloured β's)	(white β's)
	(1)	**(2)**	**(3)**	**(4)**	**(5)**	**(6)**
Actual gap in probability of advancement	**0.061**	**0.061**	**0.568**	**0.568**	**0.507**	**0.507**
I. Switching coefficients and characteristics						
Predicted gap changing all coefficients	-0.046	-0.127	0.205	-0.163	-0.022	-0.013
Percentage of gap explained	174%	307%	64%	129%	104%	102%
II. Change individual characteristics						
Change in gap due to equalizing:						
Number of grades failed	0.015	-0.036	0.038	-0.082	0.057	-0.047
Standarized LNE total score	0.062	-0.093	0.198	-0.352	0.209	-0.215
Log hh income per capita	0.023	-0.044	0.070	-0.215	0.089	-0.129
Mother's schooling	0.000	-0.008	-0.001	0.069	0.088	0.064
Father's schooling	0.015	-0.013	0.082	-0.304	0.055	-0.233
Unemployment rate	-0.019	-0.013	-0.056	0.015	0.027	0.010
Number of grades failed and LNE score	0.078	-0.123	0.243	-0.473	0.268	-0.286
Family background grouped	0.038	-0.064	0.152	-0.323	0.227	-0.195

Note: Based on probit regression in Table 4. Counterfactual simulations explained in text.

Explaining gaps in grade advancement

Table 5 uses the probits from Table 4 for counterfactuals designed to estimate the extent to which racial differences in characteristics and behaviours explain differences in grade advancement. The approach is very similar to the approach of Cameron and Heckman (2001), who look at differences in progress through school among blacks, whites, and Hispanics, in the United States. Because the coefficients are often very different between groups, we do the counterfactuals using each racial group as the baseline for any given pairwise comparison. Looking at column 1, we see that the actual gap between African and coloured students in the probability of advancing three grades between 2002 and 2005 is 6.1 percentage points. The first counterfactual assumes that Africans have the same covariates as coloureds (that is, the African coefficients are applied to the coloured sample). This counterfactual predicts a gap of -0.046, implying that Africans would have a 4.6 percentage point higher probability of advancing three grades if they had the characteristics of coloured students, given the coefficients in the African regression. In other words, we more than fully explain (we explain 174%) the gap between African and coloured students when we equalize their characteristics. In Column 2 we combine African characteristics with coloured coefficients, generating a predicted gap of -0.127, implying that African would have a 12.7 percentage point *higher* probability of advancing three grades if they had coloured characteristics. In other words, we explain 307% of the

African-coloured gap when we give the coloured coefficients to the African sample. The reason that the second counterfactual produces a larger gap in favour of Africans is the larger coefficients in the coloured regressions on variables such as the LNE scores and household income. When we assign these large coloured coefficients to the African sample, we get a large decrease in the mean predicted advancement rate, larger than the original coloured advantage.

Looking at the African-white comparisons in Column 3 of Table 5, the actual gap in the probability of advancing three grades between 2002 and 2005 is 56.8 percentage points. In the counterfactual in Column 3, which assigns African coefficients to the white sample, the predicted gap between Africans and whites drops to 20.5 percentage points. We thus explain 64% of the African-white gap when we use the African regression coefficients. Doing the counterfactual in the other direction, assigning white coefficients to Africans, we get a predicted gap of -0.163 percentage points. As was the case with the African-coloured comparison, we more than fully explain the difference in grade advancement between African and white students when we equalize their characteristics.

Recalling Figures 1 and 2, an important caveat to Table 5 is the lack of overlap in the distribution of test scores and incomes. This is most serious in the African-white comparisons, where there is virtually no overlap in the distributions of these two key variables. This means that the counterfactuals in Table 5 are largely out-of-sample projections when we assign African coefficients to

whites or white coefficients to Africans. There is nothing we can do about this lack of common support, since it is simply a manifestation of the enormous racial inequality that continues to exist in South Africa. In the case of the African and coloured comparisons the problem is less severe. The African and coloured distributions of both test scores and household income have considerable overlap. The counterfactual simulations are therefore more of a reweighting of the distribution rather than an out-of-sample projection.

The results in the top rows of Table 5 use all coefficients simultaneously, corresponding to a complete swap of characteristics between any two comparison groups. The lower part of Table 5 calculates predicted values using individual coefficients. Following Cameron and Heckman (2001), we change the value of a given characteristic for each observation in racial group X by the difference between the mean value for group X and the mean value for group Y. For example, to see the impact of giving Africans the LNE scores of coloured students, we raise the LNE score of all African students by the mean coloured LNE score advantage. We keep all other characteristics of Africans unchanged and apply the African regression coefficients to generate predicted probabilities of advancing three grades between 2002 and 2005 for each observation. Looking at column 1, we see that raising African LNE scores to the level of coloured scores and using the African regression coefficients would raise the mean predicted probability of grade advancement for Africans by 6.2 percentage points (slightly more than the total gap of 6.1 percentage

points). Column 2 shows that if we use the coloured regressions and lower coloured students to the LNE scores of Africans, their mean probability of advancement would fall by 9.3 percentage points. Column 3 shows that giving Africans the LNE scores of white students would lower the African-white gap by 19.8 percentage points (out of 56.8). Doing the exercise in the opposite direction, lowering the LNE scores of whites and using white regression coefficient, the white probability of advancement would fall by 35.2 percentage points.

Equalizing log per capita household income also has a large impact on the racial gap in grade advancement. The effect is larger when the coloured and white coefficients are used, since those coefficients are larger than the African coefficients. Column 2 shows that giving the African mean income to coloured students, using the coloured regressions, lowers the coloured probability of advancement by 4.4 percentage points, 71% of the African-coloured gap. Giving African mean income to white students, using the white regressions, lowers their probability of grade advancement by 21 percentage points, 38% of the African-white gap. The last row of Table 5 shows the impact of equalizing both the LNE scores and the number of grades behind in 2002, leaving all other characteristics unchanged. These counterfactuals show the importance of initial schooling achievement in predicting progress through secondary school. These two variables alone explain from 128% to 200% of the African-coloured gap and from 43% to 83% of the African-white gap in grade advancement.

These variables are themselves an indicator of a large number of factors that will have affected previous schooling outcomes, including school quality, household characteristics, and student's ability. While we cannot be sure exactly what caused the large racial gaps in initial test scores and grade attainment, the important point is that students entered secondary school with large pre-existing achievement gaps. Our results suggest that it would be very difficult to equalize the probability of advancing through secondary school without reducing these initial differences.

6 The impact of school type and school quality

Given persistent racial differences in school quality, it would be interesting to measure the impact of a change in school environment on progress through school. For example, it would be interesting to see whether Africans who attend predominantly white or coloured schools perform better than Africans in predominantly African schools. The empirical reality, however, is that there continues to be only limited racial mixing in schools, and very few learners change schools over the course of the four waves of the panel. This makes it hard to assess the impact of a change in school environment on the progress through school of a learner of a given race, even by using the counterfactual simulations. As was seen in Table 2, each race group had a separate education system under apartheid and many of the CAPS pupils whose progress we are analysing are still in these silos with strong differences in facilities across these school types. One major difference that does exist across learners is school fees. These are set by school governing bodies based in part on what the families of the pupils in these schools can afford to pay. As a result, school fees are strongly correlated with school type. We see very few learners who are not in these silos and even fewer who make a transition that is observed in the panel.

Table 6 assesses the impact of this situation by extending the regressions in Table 4 to include school level variables. The sample sizes are slightly smaller than those in Table 4 as we were not able to match every school named by CAPS respondents to the School Register of Needs, which is our source for school characteristics. There are a small number of schools that existed in 2000 that were not included in the School Register of Needs and some schools were created after 2000. The probit regressions in Columns 1 to 3 include the pupil-teacher ratio of the school that the young adult was enrolled in when they were interviewed in 2002. The logarithm of annual school fees in 2002 is also included in the regressions in Columns 4 and 5. We think of school fees as a proxy for a wide range of school quality variables, since schools with higher fees will be better on many dimensions. We recognize that type of school, pupil-teacher ratios, and school fees are all endogenous outcomes of the school choice decision made by students and their families. To the extent that better students choose to be in better schools, the coefficient on measures of school quality will reflect both the true effect of school quality and the correlation of school quality with unobserved ability.

Table 6 Probit regressions (marginal effects) for probability of advancing 3 grades between 2002 and 2005, CAPS respondents in Grades 8 or 9 in 2002

	Marginal Effects from Probit Regressions			Marginal Effects from Probit Regressions	
	African	Coloured	White	African	Coloured
	(1)	(2)	(3)	(4)	(5)
Grade 9 in 2002	-0.08*	-0.18***	-0.12	-0.10**	-0.17***
	[0.04]	[0.06]	[0.16]	[0.04]	[0.06]
Female	-0.03	0.04	0	-0.04	0.07
	[0.06]	[0.06]	[0.01]	[0.05]	[0.06]
Schooling performance up until 2002:					
Number of grades failed, Wave 1	-0.06*	-0.25***	-0.02	-0.06*	-0.22***
	[0.03]	[0.06]	[0.06]	[0.03]	[0.06]
Standardized LNE total score	0.11***	0.21***	0.01	0.11***	0.24***
	[0.03]	[0.05]	[0.02]	[0.03]	[0.05]
Household characteristics in 2002:					
Log hh income per cap.	0.03	0.06	0	0.02	0.06*
	[0.03]	[0.04]	[0.01]	[0.03]	[0.04]
Mother's schooling	0	0.02	0	0	0.03*
	[0.01]	[0.01]	[0.01]	[0.01]	[0.01]
Mother's schooling missing	-0.02	0.26		0.04	0.37**
	[0.13]	[0.19]		[0.13]	[0.17]
Father's schooling	0.01	0.01	0	0	0.01
	[0.01]	[0.01]	[0.00]	[0.01]	[0.01]

	(1)	(2)	(3)	(4)	(5)
Father's schooling missing	-0.02 [0.08]	-0.09 [0.12]	0 [0.00]	-0.03 [0.07]	-0.07 [0.12]
Unemployment rate					
Schooling characteristics for 2002 school:					
Learner/educator ratio (2002)	-0.01 [0.02]	0.06 [0.04]	0 [0.01]		
Log school fees (2002)	-0.01 [0.01]	-0.01 [0.01]	0 [0.00]		
Former HOA (White) school				0.44*** [0.17]	0.38*** [0.13]
Former HOR (Coloured) school				-0.15*** [0.05]	0.05 [0.08]
Constant					
Observations					

Robust standard errors in brackets in columns 1-3, p-value of F tests in parentheses in columns 4-6

* significant at 10%; ** significant at 5%; *** significant at 1%

We do not have plausible instrumental variables to identify the causal impact of school quality, so the coefficients on school characteristics should not be given a causal interpretation. While this is an important caveat to keep in mind for all of the results in Table 6, we think it is nonetheless interesting to see the extent to which school type and school quality is associated with progress through school, and to see how controlling for these variables affects the estimated impact of other individual and household characteristics.

Looking at the coefficient on the learner-educator ratio in column 2, we see that for coloured pupils there is a significant class size effect. Pupils enrolled in schools with higher pupil teacher ratios were less likely to successfully complete three grades between 2002 and 2005. For every additional pupil per teacher the probability of advancing three grades decreases by 2 percentage point. Similarly, column 5 shows that school fees have a significant effect on coloured progression. At the sample means, a one percent increase in school fees increases the probability of advancing three grades by 0.06 percentage points. However, when both these schooling variables are included (results not shown), while the size of the coefficient remains similar they are insignificant.

An important result is that introducing these school-level variables results in small changes in the impact of previous performance outcomes. The significance of the coefficient on the number of previous grades failed is reduced to the 10% level in the African regression, the coefficient on the LNE score remains the same as it was in Table 4.

Including school level variables has very little effect on the coefficients on the other variables for coloureds. Neither of the school level variables is significant in the white regression.

The fourth and fifth column of Table 6 include dummy variables for former education department. As the classification of schools created after 1994 is not clear, the small number of respondents enrolled in these schools were excluded from the regressions. The omitted category is DET (African) schools. African pupils attending a former House of Assembly (white) school have a significantly higher probability of advancing three grades than Africans in former DET schools. Surprisingly African pupils attending former House of Representatives (coloured) schools fare significantly worse than pupils in former DET schools. One would expect House of Representative schools to be much better equipped than DET schools as they would have received much more funding during the apartheid period. If one compares that characteristics of the African pupils who attend House of Representative schools with those that attend former DET schools they do appear to have better prior educational outcomes with respect to LNE scores and grades failed and come from households of higher socio-economic status. It is plausible that these pupils have a higher probability of failing as they are attending schools with higher standards. If one examines the sample of young adults who have ever been enrolled in Grade 12, Africans enrolled in House of Representative schools are significantly more likely to pass the matriculation exam. This suggests that they may ini-

tially be kept back but ultimately benefit from attending House of Representative schools. Only 7 coloured pupils attended former DET schools, so the sample in column 5 excludes these pupils. Coloureds who attend former House of Assembly schools also fare significantly better.

Table 7 presents results from counterfactual simulations that are similar to those in Table 5 but include school level variables. The coefficients for the counterfactuals are from the probits of the first three columns of Table 6. Table 7 shows the percentage of the gap that is explained when we switch coefficients in models that include school level variables and contrasts this with the percentage of the gap explained in models without school level variables. For each pairwise combination of races it is inconclusive whether the inclusion of the school level variables explains more or less of the gap in progress through school. For example, when African coefficients are applied to the coloured sample we explain less of the gap when we include the school level variables and when coloured coefficients are applied to the African sample we explain more of the gap with the school level variables. Even in the counterfactual simulations where the inclusion of school level variables allows us to explain more of the gap, the additional percentage of the gap explained is quite small. This result together with the small effects of pupil teacher ratios and school fees shown in Table 6 suggest that measures of school quality have some effect, but they explain much less of the racial differences in progress through school than variables such as the LNE scores and household income.

7 Explaining gaps in grade advancement using propensity score reweighting

The previous sections have shown that a small set of variables, including baseline test scores, household income, and parental education, can statistically account for the large racial difference in grade advancement. These results suggest that the human capital students bring with them into high school largely determines their chances of completing high school. Put another way, while there is appropriate concern about the large quality differences in high schools, our results suggest that even if African and white students were to attend identical high schools, there would still be large racial differences in grade advancement. In this section we make a final attempt to test this hypothesis using propensity score matching. We assess whether a racial gap in the probability of progressing three grades in three years remains when students are matched on the human capital content they bring to high school in addition to the characteristics of their high school.

The matching approach can be thought about in the following way: if race could be randomly assigned to respondents in 2002, what would the effect of race be on grade progression in the next 3 years? The policy question is therefore what the effect on secondary school grade progression would be if the distribution of primary school fees, learner teacher ratios, household incomes parental education levels and primary school performance were equivalent across racial groups.

Table 7 Predicted change in the racial gap in grade advancement from changing coefficients across samples and from changing values of individual coefficients.

	African equated to coloured	Coloured equated to African	African equated to white	White equated to African	Coloured equated to white	White equated to coloured
	(African β's)	(coloured β's)	(African β's)	(white β's)	(coloured β's)	(white β's)
	(1)	**(2)**	**(3)**	**(4)**	**(5)**	**(6)**
Actual gap in probability of advancement	**0.061**	**0.061**	**0.570**	**0.570**	**0.509**	**0.509**
I. Switching coefficients and characteristics						
Predicted gap changing all coefficients	-0.038	-0.134	0.176	-0.079	-0.048	0.088
Percentage of gap explained	162%	318%	69%	114%	109%	83%
II. Change individual characteristics						
Change in gap due to equalizing:						
Number of grades failed	0.018	-0.054	0.033	-0.173	0.044	-0.065
Standarized LNE total score	0.063	-0.081	0.195	-0.314	0.168	-0.185
Log hh income per capita	0.027	-0.031	0.078	-0.295	0.057	-0.169
Mother's schooling	-0.001	-0.008	-0.008	0.064	0.063	0.058
Father's schooling	0.016	-0.007	0.081	-0.232	0.028	-0.174
Unemployment rate	-0.022	-0.010	-0.070	0.051	0.023	0.037
Log school fees (2002)	-0.006	-0.033	-0.026	0.075	0.120	0.061
Learner educator ratio (2002)	0.016	-0.016	0.069	-0.002	0.050	-0.001
Number of grades failed and LNE score	0.083	-0.128	0.233	-0.542	0.213	-0.285
Family background grouped	0.045	-0.047	0.164	-0.400	0.155	-0.234
School characteristics grouped	0.010	-0.048	0.040	0.074	0.173	0.060

Note. Based on probit regression in Table 6. Counterfactual simulations explained in text.

Table 8a Internal balancing checks

| | African (452) - Coloured (385) | | | | | Coloured (385) - White (83) | | | | |
| | Original | | | Reweighted | | Original | | | Reweighted | |
	Diff. in means	T-stat	P-value	T-stat	P-value	Diff. in means	T-stat	P-value	T-stat	P-value
Grade 9 in 2002	0.01	0.36		0.65		0.04	0.53		-0.53	
Number of grades failed by 2002	0.22	3.30	**	-1.87		0.34	4.48	***	0.38	
Standardized LNE total score	-0.59	-8.54	***	-0.38		-1.11	-11.32	***	-1.46	
Log per cap hh income (mean zero)	-0.89	-10.26	***	0.47		-1.66	-13.99	***	-0.87	
Mother's schooling	-0.49	-1.65		0.63		-3.63	-13.10	***	-2.14	*
Mother's schooling missing	0.01	0.48		-0.49						
Father's schooling	-1.62	-4.26	***	0.23		-5.92	-9.07	***	-0.37	
Father's schooling missing	0.08	2.06	*	-0.72		0.24	4.60	***	-0.21	
Annual school expenses in 2002 (rands)	-0.53	-3.60	***	0.92		-2.52	-7.87	***	-0.03	
Annual school expenses in 2002 missing	-0.01	-0.75		0.51		0.05	2.24	*	2.86	**
Pupil-teacher ratio in 2002	0.85	0.91		0.27		7.23	4.80	***	0.52	
Pupil-teacher ratio in 2002 missing	0.03	0.92		-0.15		-0.07	-1.39		0.08	

Table 8b External balancing checks

| | African (452) - Coloured (385) | | | | | Coloured (385) - White (83) | | | | |
| | | Original | | Reweighted | | | Original | | Reweighted | |
	Diff. in means	T-stat	P-value	T-stat	P-value	Diff. in means	T-stat	P-value	T-stat	P-value
Connected to electricity	-0.09	-3.05	**	-2.23	*	-0.02	-1.53		-1.69	
Radio	-0.08	-2.66	**	0.49		-0.13	-6.42	***	0.24	
TV	-0.15	-4.52	***	-0.65		-0.07	-3.09	**	-4.56	***
Fridge	-0.26	-7.25	***	-4.34	***	-0.04	-1.70		-3.63	***
More than 5 books	-0.22	-6.62	***	-1.71		-0.11	-5.77	***	-5.77	***
Telephone (not cellular)	-0.35	-8.46	***	-2.59	*	-0.27	-5.28	***	1.01	
Stive	-0.36	-11.29	***	-10.85	***	0.02	1.13		1.35	
Car, Bakki or Combi	-0.31	-8.20	***	-1.92		-0.55	-15.67	***	-2.17	*
Internal piped water	-0.54	-10.53	***	-8.88	***	-0.08	-4.41	***	-4.41	***
Washing machine	-0.67	-21.26	***	-6.52	***	-0.20	-5.88	***	0.01	
Permanent building	-0.40	-8.21	***	-8.26	***	-0.06	-1.96		-5.22	***
Former DET (African) school	0.73	24.87	***	27.12	***	-0.11	-5.77	***	-5.77	***
Former HOA (White) school	-0.04	-2.22	*	2.13	*	-0.07	-0.84		-3.48	***
Former HOR (Coloured) school	-0.65	-20.82	***	-14.61	***	0.02	1.82		1.82	
New school since 1994	0.07	4.19	***	4.67	***	-0.78	-16.90	***	-7.82	***

Propensity score matching has become popular in the economics literature to control for economic endogeneity and selection or omitted variable bias. In our context, we use the method to address selection at the time students start secondary school since by this time students have accumulated different levels of disadvantage by race most of which affect progress through secondary school. Generating a plausible counterfactual is therefore the main challenge in assessing the effect of race on progress through secondary school. This entails creating a counterfactual group which has very similar characteristics in 2002 to their comparison race, except that they differ by race. As shown in Table 2 and Figures 1 and 2, apartheid has left large racial differences in attainment by secondary school, household income and parental education levels and school quality. We therefore estimate the probability of being a certain race using these characteristics and use this propensity score to create a counterfactual distribution from the racial comparison group.

The propensity score method relies on two important assumptions; the conditional independence assumption (CIA) and the common support assumption. The plausibility of the CIA assumption rests on the data; we require data that can account for selection into being a certain race, which requires that the data be rich enough to account for differential levels of accumulated disadvantage in 2002. Tables 8a and 8b present tests of the difference in the means of the observable characteristics between Africans and coloureds in the left panel and coloureds and whites in the right panel. The aim of propensity

score reweighting is to reweight the distribution of the observable characteristics in the population group of interest to match the distribution in another group. Thus if the reweighted means are not significantly different between the two groups of comparison, the data are balanced and the CIA is achieved.

Table 8a presents internal balancing tests; variables used in the construction of the propensity score are compared once the data are reweighted. We see, for example, in the second row of table 8a, that Africans had failed significantly more grades by 2002 than coloureds. However, once the data are reweighted the difference between means is insignificant. By design, the propensity score attempts to minimize the difference between racial groups for the variables included in the propensity score estimation. Thus a more stringent balancing test is applied in Table 8b, external balancing tests. Here we check for balance between racial groups on variables not included in the propensity score estimation. We find that while the difference between the mean levels between racial groups decreases in most cases, many of these differences remain highly significant.

One advantage of the matching method over running regressions is its ability to attend, to an extent, to the common support issue. As noted in previous sections, characteristics between races, especially Africans and whites have limited overlap. In the previous models, counterfactuals for Africans whose characteristics were outside the common support region were derived solely from projections based on the specific functional form of

the model. Matching estimators put restrictions on the joint distribution to aid the identification assumption. More specifically, we use a kernel matching estimator which assigns differential weights to comparison observations based on the estimated propensity score thus determining which observations are used, and their importance, in the constructed counterfactual. In other words, the common support assumption requires that the positive support regions for the observable dis-

tributions should be identical for different races. While the common support assumption is a benefit of the matching estimator, given the lack of overlap of African and white characteristics for almost all predictive observable characteristics, the method can only be applied for comparisons between African and coloured and coloured and white. Figures 4 and 5 confirm graphically that the assumption of common support is met in both these instances.
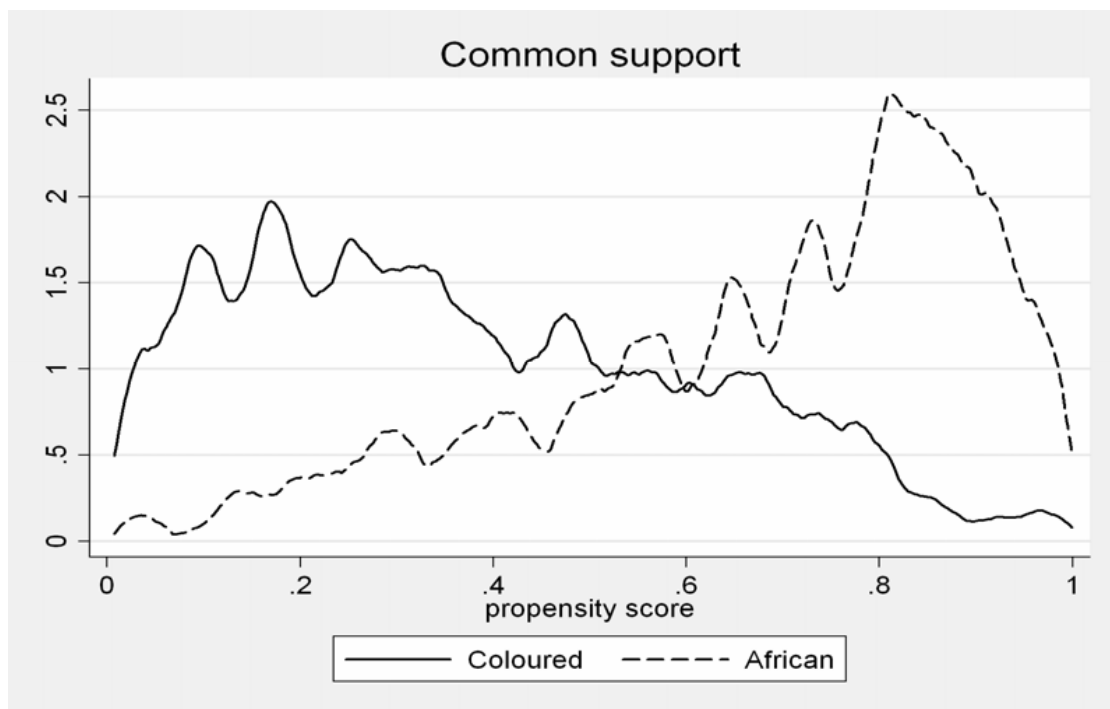

Figure 4

Table 9 presents estimates of the impact of being African versus coloured on progression through secondary school in the left hand panel and similar estimates for the comparison between coloureds and whites in the right hand panel. The first column of table 9 (Naïve) presents the unconditional dif-

ference in the average level of completing three grades by 2005 for Africans versus coloureds. Africans are 7 percentage points less likely to complete three years of secondary school than coloureds between 2002 and 2005. The second column presents the marginal effect of being African when the

variables used to calculate the propensity score are included as a set of controls in a probit regression of progressing on an African indicator. Inclusion of these controls switches the sign on the African coefficient – Africans are 17 percent more likely to progress three grades by 2005 than coloureds.

The third column presents the matched estimate, the impact of being African when comparison is made between African and coloureds of similar baseline characteristics. Africans are 15 percent more likely to progress three grades than coloureds.
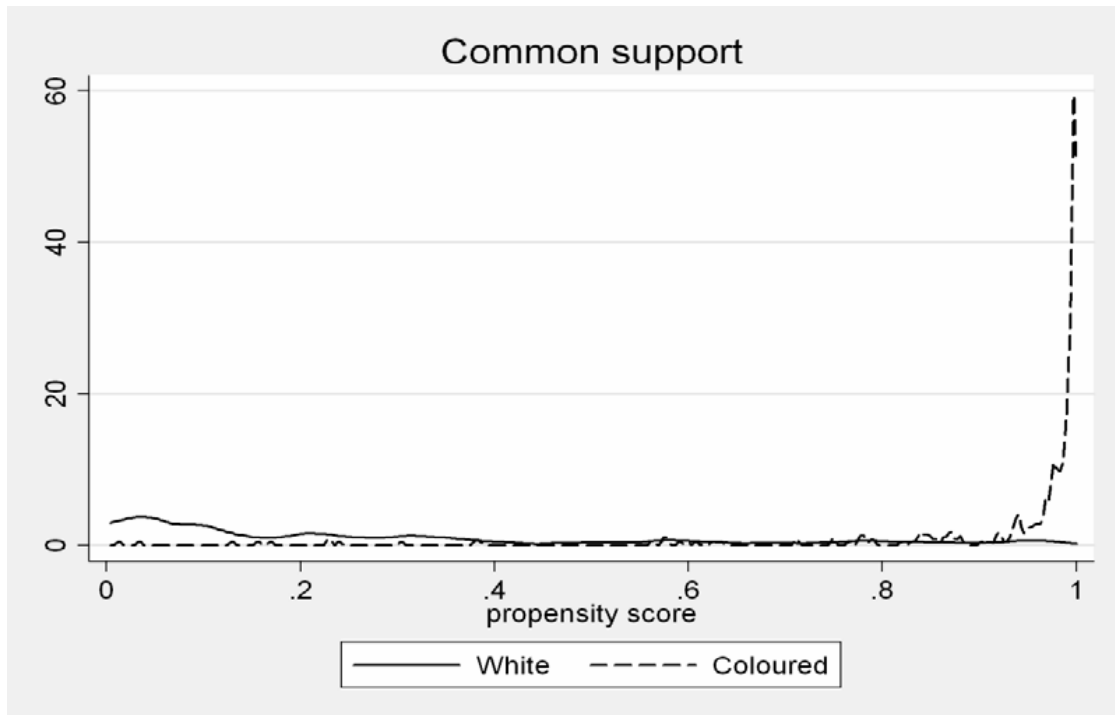

Figure 5

Shifting to the right hand panel of Table 9, we see large difference in the unconditional mean of progressing three grades between coloureds and whites. However, we find no significant impact of being coloured versus white on the chance of progressing using the matching estimator. This signifies that the progression gap between coloured and whites is completely explained by the baseline characteristics. This is consistent with our results using regression to control for baseline characteristics.

8 Conclusions

High levels of grade repetition play a fundamental role in explaining the persistent racial gaps in schooling attainment in South Africa. Using recently collected longitudinal data from the Cape Area Panel Study in Cape Town, we find large racial differences in the probability that 8th and 9th Graders make normal progress through secondary school between 2002 and 2005. While 82% of white students in the 8th

and 9th Grade in 2002 had reached Grade 11 or Grade 12, respectively, by 2005, only 27% of African students and 34% of coloured students had made the same progress. Although grade repetition and dropping out of school both contribute to the gap in grade advancement, grade repetition is by far the most important factor, especially for Africans.

The results of our probit regressions indicate that per capita household income and indicators of previous achievement such as test scores and the number of grades behind in 2002 are strong predictors of subsequent progress through school. Estimating counterfactuals using our separate probit regressions for each race, we find that we can entirely explain the racial gaps in grade achievement between 2002 and 2005 by differences in the characteristics of students and their households in 2002. Taken at face value, the results suggest that eliminating the large racial differences in the quality of secondary schools would have limited impact on the racial gap in grade progression. We explore this directly by including a set of school level variables in the probit regressions and in the counterfactual analysis. Individually and as a set, these variables do not have a significant impact on grade advancement. Moreover, their inclusion in the models lowers the impact of student and household characteristics only marginally.

Table 9 Marginal effect of race on the probability of progressing 3 grades in 3 years from Naïve (unconditional) difference in means, probit with controls and kernel propensity score reweighting.

	African compared to Coloured			Coloured compared to White		
	Naïve	Probit	Kernel	Naïve	Probit	Kernel
ref: Coloured				*ref: White*		
African	-0.07*	0.17***	0.15***	**Coloured**	-0.48***	0.07	-0.08
	[0.04]	[0.04]	[0.03]		[0.05]	[0.11]	[0.20]
Observations	837	837	837	Observations	468	432	466

Robust standard errors in brackets

* significant at 10%; ** significant at 5%; *** significant at 1%

African students begin secondary school with such large disadvantages in terms of test scores and previous school performance that it is very difficult for them to complete school at the same rate as coloured or white students. When they were given the test scores and previous failure rates of coloured and white students in the simulations, they performed at least as well as these students performed. It seems then that while the demonstrably inferior secondary schools in predominantly African areas can and do perpetuate the poor performance of African students, they are not the primary cause of this poor performance. The root of these schooling problems lies much earlier in the poor quality primary schools and the disadvantages associated with growing up in poor households. The important policy challenge is to give particular attention to these pre-secondary school spheres along with a commitment to improving the quality of the secondary school themselves.

References

Anderson, Kermyt G., Anne Case, and David Lam (2001) "Causes and Consequences of Schooling Outcomes in South Africa: Evidence from Survey Data." *Social Dynamics* 27(1): 1-23.

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

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

Case, Anne, and Angus Deaton (1999)

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

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

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

Lam, David, Ardington, Cally, Branson, Nicola, Case, Anne, Leibbrandt, Murray, Menendez, Alicia, Seekings, Jeremy, Sparks, Meredith (2008). *The Cape Area Panel Study: Overview and Technical Documentation of Waves 1-2-3-4*. The University of Cape Town, October 2008.

Lam, David, Cally Ardington, and Murray Leibbrandt (2011) "Schooling as a Lottery: Racial Differences in School Advancement in Urban South Africa," *Journal of Development Economics* 95: 121-136.

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

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

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