
Quantitative and qualitative aspects of education in South
Africa:
An analysis using the National Income Dynamic Study

MIA DE VOS

Stellenbosch Economic Working Papers: 06/11

KEYWORDS: EDUCATION QUALITY, COGNITIVE SKILLS, SOUTH AFRICA, RACIAL
DIFFERENTIAL
JEL: I21

MIA DE VOS
DEPARTMENT OF ECONOMICS
UNIVERSITY OF STELLENBOSCH
PRIVATE BAG X1, 7602
MATIELAND, SOUTH AFRICA
E-MAIL: 14576848@SUN.AC.ZA



UNIVERSITEIT
STELLENBOSCH
UNIVERSITY



A WORKING PAPER OF THE DEPARTMENT OF ECONOMICS AND THE
BUREAU FOR ECONOMIC RESEARCH AT THE UNIVERSITY OF STELLENBOSCH

Quantitative and qualitative aspects of education in South Africa: An analysis using the National Income Dynamic Study¹

MIA DE VOS

ABSTRACT

Based on a graphical and statistical analysis of the National Income Dynamic Study (NIDS), this paper provides a comprehensive picture of the educational context in South Africa. The main question under consideration is whether quantitative and qualitative educational attainment differs significantly along racial lines. The data shows that the government has been largely successful in reducing the race-based educational gap in terms of school enrolment and years of education successfully completed. Matriculation results and numeracy test scores unfortunately suggest that higher levels of educational attainment do not necessarily reflect positively on educational outcomes. This implies that the South African educational system is still characterized by large differentials in the quality of education.

Keywords: Education quality, Cognitive skills, South Africa, Racial differential
JEL codes: I21

¹ This document was produced within the Social Policy Research Group in the Department of Economics at Stellenbosch University with the financial assistance of the PSPPD (Programme to Support Pro-Poor Policy Development in South Africa), a partnership programme between The Presidency, Republic of South Africa, and the European Union (EU). The contents do not reflect the position of The Presidency or the EU.

1. Introduction

Traditionally an individual's level of education has been measured by the discreet years of schooling successfully completed by such individual (Chamberlain & Van der Berg, 2002: 1-2). The difficulty with this common measure of educational attainment, however, is that it can be very misleading since it assumes that years of schooling is an accurate indication of the effective level of education attained (Hanushek, 2007: 2; Chamberlain & Van der Berg, 2002: 1-2). Most people would agree that a year of schooling in a rural village under a tree does not produce the same level of cognitive skill and knowledge as a year in a private school in an urban area (Hanushek & Woessmann, 2008: 608). This is especially true in a South African context, where time spent in school does not necessarily provide a clear picture of effective years of schooling completed – mainly due to differences in the quality of education received (Chamberlain & Van der Berg, 2002: 1-2). According to Van der Berg (2007: 6) promotion to the next level of education is furthermore relatively easy, leading to the conclusion that years of education completed may overstate progress in cognitive skills mastered. Nevertheless, and unfortunately, the majority of education economics literature ignores these vital issues (Hanushek & Woessmann, 2008: 608).

This paper aims to provide a comprehensive picture of the educational context in South Africa by investigating quantitative and qualitative educational differentials by race. In particular this implies that educational attainment should be measured in terms of years of education completed as well as level of cognitive skill actually acquired. The numeracy test scores provided by the NIDS dataset combined with school quality data allows one to explore these differentials in cognitive skill.

The author commences the paper with a description of the dataset used in addressing this topic, including a discussion on the difficulties associated with this data. In section 3 various statistical and analytical techniques are utilised in order to analyse the educational context in South Africa. Section 3.1 provides a detailed description of historical trends in quantitative educational attainment in South Africa. Section 3.2 contains an analysis of school quality data and the numeracy test scores in order to promote an understanding of the qualitative educational trends. In section 3.3, regressions are executed using numeracy test score as dependent variable. Section 4 provides some concluding comments.

2. Data and Methodology

2.1 Data

This paper is based on an analysis of the first national panel study in South Africa, the National Income Dynamic Study (NIDS). The South African Presidency initiated and funded this multi-million Rand survey in an attempt to follow changes in the welfare of South Africans over time, and NIDS forms part of the Presidency's Programme to Support Pro-Poor Policy Development (PSPPD). The first Wave of NIDS was conducted in 2008 by the South African Labour and Development Research Unit (SALDRU) at the University of Cape Town's School of Economics. Approximately 300 fieldworkers were sent across the country to collect information on and from 7305 households that consist of 28 000 individuals in total. Every two years field workers will go back to the same respondents that they visited for the first Wave. Although the second Wave of data collection was carried out in 2010; the results have not yet been published (Leibbrandt, Woolard & de Villiers, 2009: 1).

The authors' main consideration for using the NIDS dataset is that it contains scores on a numeracy test taken by a sub-sample of the survey population. Given that performance test scores are determined by not just inherent ability but also by quality of education received, data on numeracy test scores is invaluable in promoting an understanding of variations in school quality in South Africa (Van der Berg, Wood & Le Roux, 2002: 290; Hanushek & Woessman, 2008: 608; Van Broekhuizen & Von Fintel, 2010: 4).

Respondents between the ages of 15 and 59 had the option of writing a 10 minute numeracy test. The test was designed to evaluate a broad range of mathematical domains related to the South African school curriculum. Tests consisted of 15 multiple choice questions, including, *inter alia*, basic calculations; geometry; algebraic expressions; basic probability theory; and conversion between time, money and volume units. In order to control for the fact that respondents differed in terms of educational attainment, four different test levels were made available. Consequently, respondents had to write the test that corresponded with the highest level of mathematics they had attained. For example, respondent with no education and up to grade 3 qualified for the first test level, *et cetera*. Each level contained questions regarding mathematics skills considered to be taught at that level.

The NIDS dataset had the added advantage of providing the names of the schools that respondents went to, and the schools were categorised according to former racially-based departments. The matriculation results of all the schools in the dataset were furthermore obtained for 2004 and 2005. Each school was assigned a set of scores for maths higher grade (HG) and standard grade (SG) and English first and second language, based on the average performance in those subjects of all children that attended that particular school.

In addition to scores on the above-mentioned subjects, an overall performance score calculated using RASCH analysis was given to each school². It should be emphasised that the dataset does not provide respondents' individual scores, but rather an average score for the school that respondents attended. Using the output-based approach, this ultimately allows researchers to draw conclusions about the variation in quality of the education provided by different schools. Unfortunately spelling errors, changes in school names and missing information made it extremely difficult to do the pairing.

The NIDS dataset also provides data on a range of labour market outcomes such as employment and earnings, which allows researchers to investigate the role that poor educational quality has to play in explaining labour market inequalities (Van Broekhuizen & Von Fintel, 2010: 4; Hanushek & Woessman, 2008: 616).

2.2 Data Limitations

For the purposes of evaluating and monitoring South Africa's educational system, it is informative to analyse the NIDS dataset in order to identify quantitative and especially qualitative trends. Cognisance should, however, be taken of some data limitations that might jeopardise the precision and statistical power of the results.

A serious concern regarding the data being used is the relatively small sample sizes, especially amongst coloured and Indian individuals. The data suggests that the surveyed sample is not always a true reflection of the target population. This can be ascribed to the fact that the questions in the survey were answered on a voluntarily basis. In order to correct for the over- or under-sampling of certain groups, weights are used in the analysis. As a result of the data limitations, the author

² The performance scores were estimated by Cobus Burger from the University of Stellenbosch in collaboration with SALDRU.

focuses on the white and African population groups for the majority of the results presented in the rest of the paper.

Considering that the numeracy test was also taken on a voluntarily basis, the author was posed with the significant challenge of sample selection bias. The top panel of Table 1 shows that there is a definite racial bias in the numeracy test response rates, while the bottom panel displays that the younger cohorts were more willing to write the test than their older cohorts. The total column furthermore illustrates that the overall response rate was extremely low. In a recent paper, Van Broekhuizen & Von Fintel (2010) extensively analyse the low and differential response rate in the NIDS numeracy test and possible reasons for it.

Table 1: Individuals Who Wrote the Numeracy Test, by Race and Age Cohort

	No	% No	Yes	% Yes	Total
Race					
African	9326	76.73%	2829	23.27%	12155
Coloured	1916	78.20%	534	21.80%	2450
Indian	254	93.38%	18	6.62%	272
White	748	86.57%	116	13.43%	864
Total	12244	77.78%	3497	22.22%	15741
Age Cohorts					
15-19	1825	58.87%	1275	41.13%	3100
20-24	1816	71.52%	723	28.48%	2539
25-29	1488	76.58%	455	23.42%	1943
30-34	1334	80.22%	329	19.78%	1663
35-39	1302	83.84%	251	16.16%	1553
40-44	1305	88.84%	164	11.16%	1469
45-49	1172	88.39%	154	11.61%	1326
50-54	1068	91.36%	101	8.64%	1169
55-59	975	95.49%	46	4.51%	1021
50-54	1068	91.36%	101	8.64%	1169
55-59	975	95.49%	46	4.51%	1021
Total	14328	79.72%	3645	20.28%	17973

Source: Own calculations using NIDS (2008)

It appears furthermore as though there were problems with the recording of the numeracy test score data. There were four different test levels, and respondents were supposed to take the level that corresponds with the highest level of mathematics attained in school. However, the data suggests that field workers allowed respondents to take any level they felt comfortable with, i.e. in some instances persons with a bachelors degree wrote the level 1 test or a person with no

mathematics education completed the highest test level. This implies that the test score data are not a very reliable reflection of the subsample that took the test.

The issue of whether a numeracy test is actually a credible indicator of cognitive ability is also important. Since the tests included a broad range of mathematical domains³, more recent exposure to education may have had a positive impact on test scores. In other words, if a 40 year old medical doctor performed poorly in the numeracy test, it does not necessarily imply that the person has a low level of cognitive ability, but rather that this person’s basic mathematical skills may no longer be so sharp.

3. Results

3.1 Quantitative Educational Attainment

Table 2: Mean Years of Education Completed⁴ by Race and Age

	African			Coloured			Indian			White			Total		
	Mean	St dev	CV	Mean	St dev	CV	Mean	St dev	CV	Mean	St dev	CV	Mean	St dev	CV
	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15
15-19	9.19	1.96	0.21	9.34	1.77	0.19	9.66	1.63	0.17	10.36	1.70	0.16	9.29	1.95	0.21
20-24	10.40	2.36	0.23	10.32	2.36	0.23	11.84	2.56	0.22	11.73	1.92	0.16	10.51	2.37	0.23
25-29	10.20	2.78	0.27	10.53	2.57	0.24	11.46	2.46	0.22	12.31	2.31	0.19	10.40	2.78	0.27
30-34	9.74	3.43	0.35	10.63	3.19	0.30	12.13	2.49	0.21	12.26	1.76	0.14	10.07	3.38	0.34
35-39	8.76	3.99	0.46	9.93	3.16	0.32	10.95	2.15	0.20	12.35	1.60	0.13	9.31	3.85	0.41
40-44	7.81	4.11	0.53	8.23	3.59	0.44	11.11	2.22	0.20	12.43	2.19	0.18	8.64	4.12	0.48
45-49	6.80	4.33	0.64	8.53	3.62	0.43	10.03	4.21	0.42	13.41	2.52	0.19	8.15	4.68	0.57
50-54	5.91	4.20	0.71	6.56	4.09	0.62	9.87	3.93	0.40	12.33	2.29	0.19	7.20	4.60	0.64
55-59	4.43	3.86	0.87	6.52	4.18	0.64	7.22	4.25	0.59	12.75	2.72	0.21	6.38	4.92	0.77
60-64	4.22	3.95	0.94	4.33	4.32	1.00	6.47	3.77	0.58	11.34	2.24	0.20	5.93	4.70	0.79
65-69	3.16	3.85	1.22	6.07	3.98	0.66	6.00	4.37	0.73	11.95	2.79	0.23	5.55	5.16	0.93
70-74	2.63	3.60	1.37	6.09	3.08	0.51	2.54	2.77	1.09	11.90	2.61	0.22	5.12	5.12	1.00

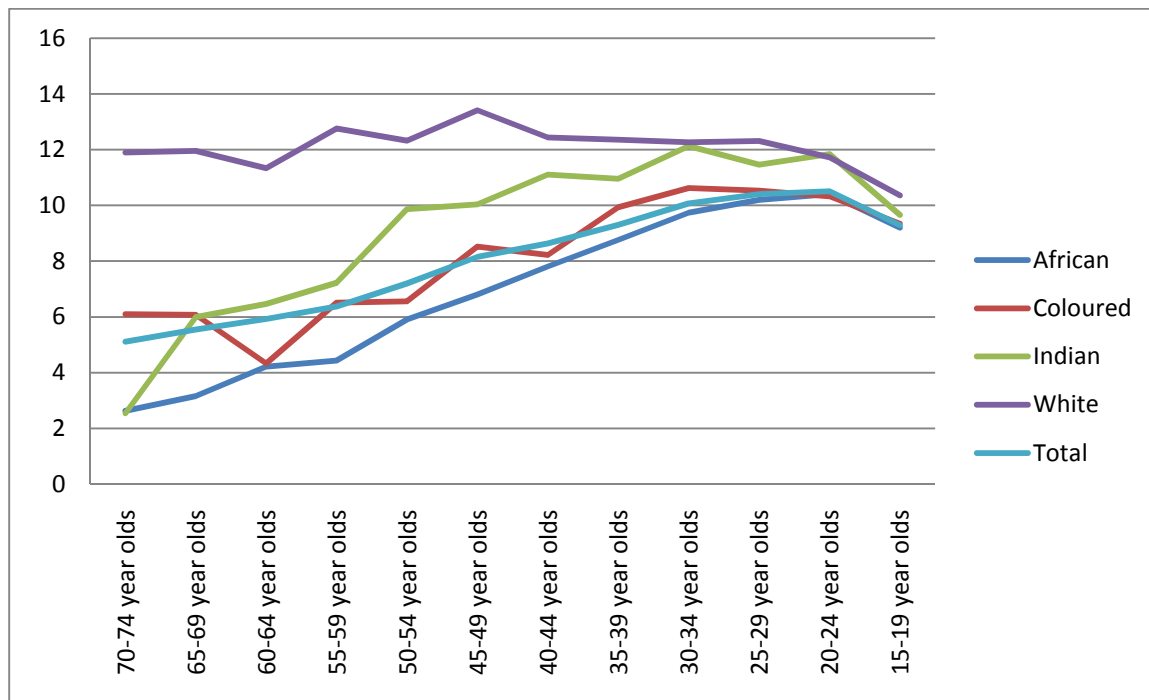
Source: Own calculation from NIDS 2008

Table 2 presents descriptive statistics for the quantitative educational attainment across different race groups for five-year age cohorts in 2008. This table provides a broad overview of the history of education in South Africa from cohorts born in the late 1930s to cohorts born in the early 1990s. Three measures are used: mean years of education completed, standard deviation in years of schooling and coefficient of variation in years of schooling.

³ See http://www.nids.uct.ac.za/home/index.php?option=com_docman&task=cat_view&gid=12&Itemid=19 for an example of the numeracy tests.

⁴ Years of education completed refers to years completed successfully, i.e. highest grade completed and does not necessarily correspond to the actual number of years individuals spent in the school system.

Figure 1: Mean Years of Education Completed by Race and Age Cohort



Source: Own calculations from NIDS 2008

Figure 1 graphically shows that racial inequalities in terms of years of schooling completed decreased substantially over the years. The greatest distinction in mean years of education completed can be seen between the African and white populations of the 70-74 year old cohort (born between 1952 and 1956). African individuals had on average attained 9.3 fewer years of education than their white counterparts. The low levels of quantitative educational attainment amongst the older Africans can be ascribed to the fact that these cohorts have been educated during the apartheid years, when educational resources were mainly allocated to so-called white schools (Chamberlain, 2001: 9; Van der Berg, 2007: 3). It should be pointed out that the racial differential had been reduced even during the apartheid era; for instance, for the 40-44 year age group (born between 1964 and 1968) the African-white gap fell to 4.62. Over the years this racial difference continued its downward trend and dropped to 1.17 years for the 15-19 year old cohort (born between 1989 and 1993).

It can also be seen that the coloured and Indian population experienced a dramatic improvement in terms of educational attainment – mean years of education completed gradually increased over the years. The Indian population shows higher levels of educational attainment than their coloured and African counterparts throughout, with the exception of the two oldest cohorts. This finding is supported by Chamberlain (2001: 16), who states that during the previous educational system, the Indian education department performed remarkably well despite deficient resources. One should

furthermore point out how the African and total population curves converge as one moves from the oldest to the youngest cohort – an illustration of how the African population is rapidly becoming a larger proportion of South Africa’s total population (Louw *et al.*, 2006: 12).

The standard deviation in years of education completed is also displayed in Table 2. Lam (1999: 7) shows that this summary statistic is an important indicator of earnings inequality amongst race groups.

When explaining the link between earnings inequality and standard deviation in years of schooling, Lam (1999: 3-4) uses the standard earnings function as a point of departure. If one omits experience and other determinants of earnings for the moment, the logarithm of the *i*th worker’s earnings is

$$\log y_i = \alpha + \beta S_i + u_i \dots \dots (1)$$

(where y_i is earnings, S_i is schooling and u_i is a residual uncorrelated with schooling).

Consequently, the variance of log earnings is:

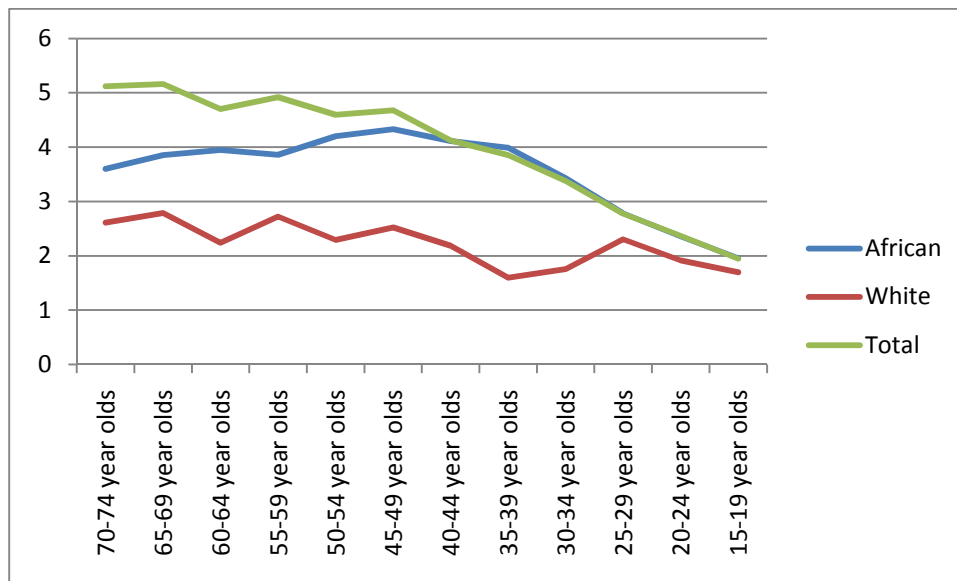
$$\text{var}(\log y_i) = \beta^2 \text{var}(S_i) + \text{var}(u_i) \dots \dots (2)$$

(where $\text{var}(\log y_i)$ is a standard mean-invariant measure of earnings inequality). Equation 2 illustrates that if the link between earnings and schooling is log-linear as shown in equation 1, then earnings inequality ($\text{var}(\log y_i)$) is linearly-related to the variance (standard variation squared) in schooling (Lam, 1999: 3-4).

Figure 2 displays some interesting patterns. The standard deviation for the African population group rises steadily for about 30 years, reaches a maximum in the 45-49 year old cohort, and then drops substantially with each cohort as one move towards the youngest cohort (see column 2 in Table2). This implies that earnings inequality within the African race group increases for each cohort from the oldest cohort to the 45-49 year old cohort, and then declines steadily towards the youngest cohort. A study by Lam (1999) using the 1995 October Household Survey displays a similar pattern.

As one would expect, the standard deviation for the white population group (see column 11 in Table 2) is lower than that of the African population at all ages; and also shows a gradual decline from the oldest cohort to the youngest. Column 14 shows that the standard deviation for the total population, a value that is important in explaining overall earnings inequality in South Africa, drops significantly from 5.12 in the oldest cohort to 1.95 in the youngest cohort.

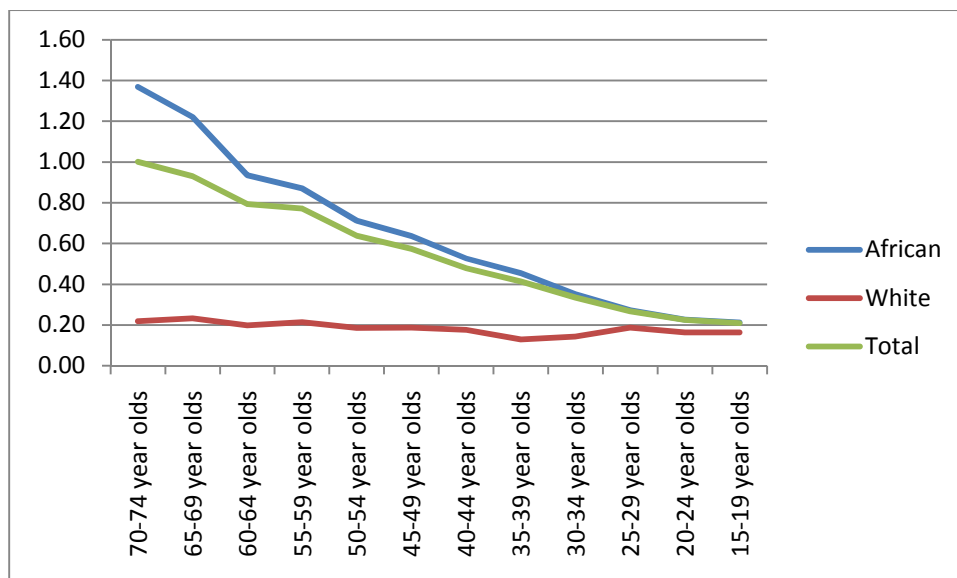
Figure 2: Standard Deviation in Years of Education Completed



Source: Own calculations from NIDS 2008

Table 2 further shows the coefficient of variation in years of education completed. According to Lam (1999: 7), the coefficient of variation, $CV = \frac{\sigma}{\mu}$ (standard deviation divided by mean) is a standard mean-invariant measure of intra-racial schooling inequality.

Figure 3: Coefficient of Variation in Years of Education Completed



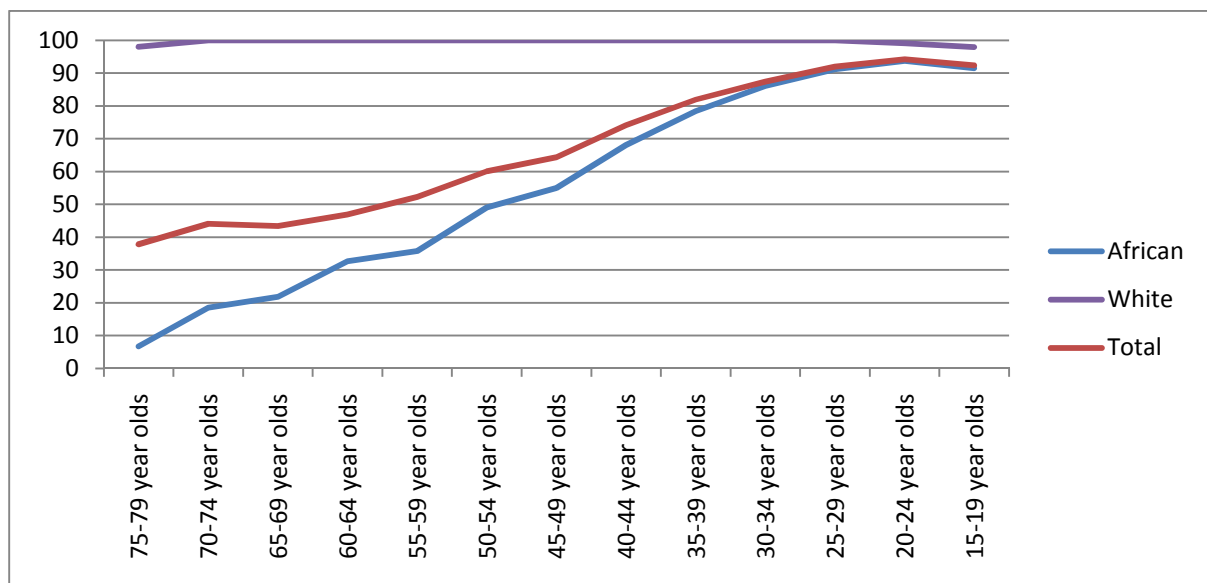
Source: Own calculations from NIDS 2008

Figure 3 graphically indicates steady declines from the oldest to the youngest cohorts for the African race group. The coefficient of variation for the white population is considerably lower than that of

their African counterparts for the older cohorts. This gap steadily declines, however, as one move towards the youngest cohort. The coefficient of variance for the total South African population furthermore drops substantially as one move from left to right in Figure 3 – a trend that suggests that there was a substantial decline in schooling inequality among all race groups over the 60-year period.

Figure 4 & 5 show the proportions of South African learners that have successfully completed primary and secondary school. See Appendix A for the tables from which these figures were obtained. In Figure 4 it can be seen that the white population has almost universal completion of grade 7 for the past six decades. For the older cohorts, African individuals lagged considerably behind their white counterparts, but Figure 4 graphically shows how this gap has narrowed noticeably as one move towards the younger cohorts. The total percentage of South Africans completing primary school furthermore increased from 37.82% for the 75-79 year old cohort to 92.3% for the 15-19 year cohort.

Figure 4: Proportion Completing Grade 7, African & White Population Groups



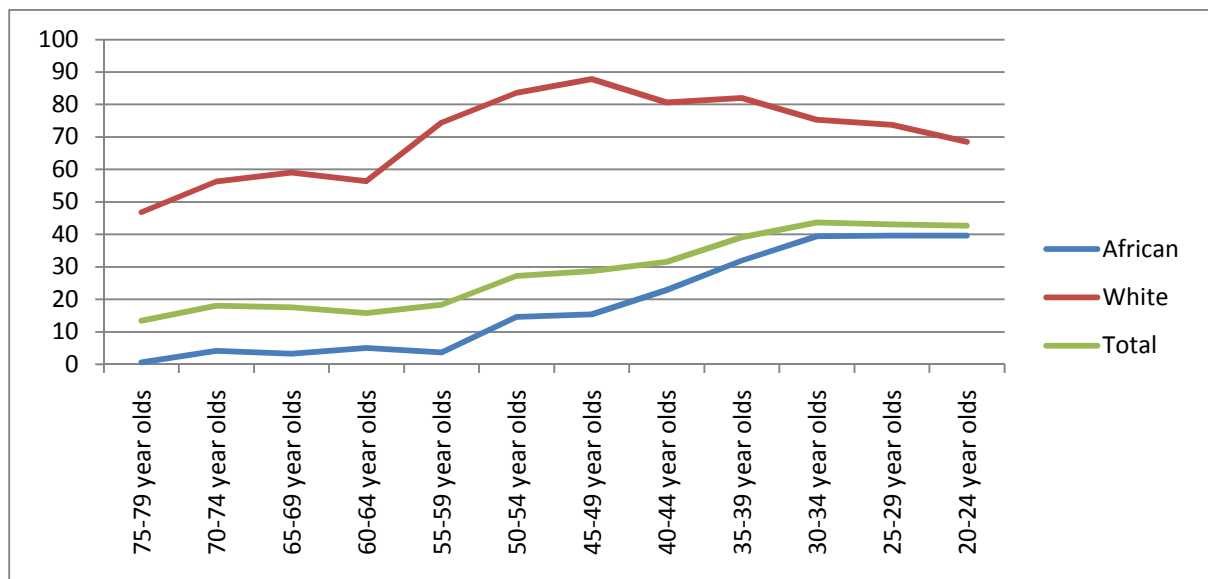
Source: Own calculations from NIDS 2008

Figure 5 demonstrates some alarming trends. The progress that has been made in terms of primary school completion cannot yet be seen in completion of secondary school. Although the proportion of African pupils with at least grade 12 has increased dramatically from 0.64% for the oldest to 39.61% for the youngest cohort, there still exists a large gap between African pupils and their white counterparts. Lam (1999: 8) describes this phenomenon as follows: “The least progress in closing the

gap between whites and non-whites in South Africa has been made in completion of secondary school.”

Figure 5 furthermore shows that the proportion of white pupils completing grade 12 has declined in recent years; the proportion of white pupils with at least grade 12 peaks in the 45-49 year old cohort and then drops steadily to 68.51% for the youngest cohort.

Figure 5: Proportion Completing Grade 12, African and White Population Groups



Source: Own calculations from NIDS 2008

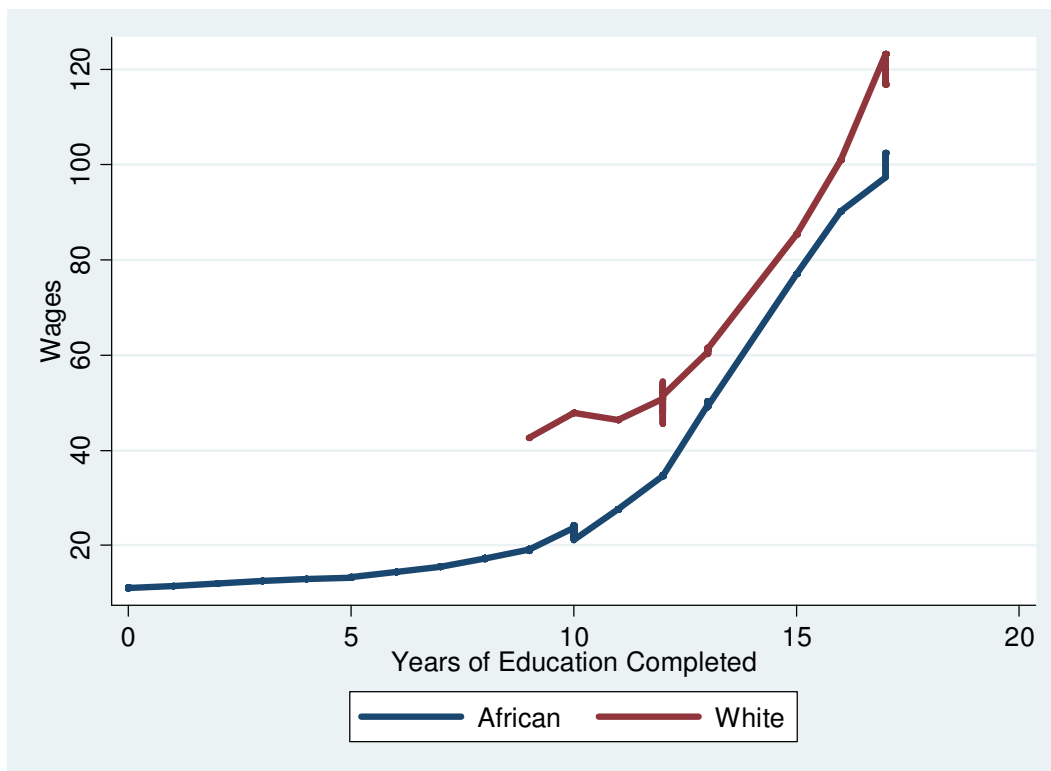
These trends are disturbing since an individual’s opportunities for further education as well as his or her chances in the labour market depends to a great extent on the completion of grade 12 (Anderson *et al.*, 2001: 5).

Furthermore, there has not been much improvement in terms of the number of African university graduates. Only 1.4% of Africans above the age of 26 have a degree, in comparison to almost 15% of whites (Van der Berg, 2007: 4). Van der Berg (2007: 4) states that this phenomenon can mainly be ascribed to poor school quality.

Historically, large earnings inequalities have been ascribed to substantial educational differentials along racial lines. These descriptive statistics demonstrate that the government has been largely successful in reducing quantitative educational attainment differentials – not just *between* race groups but also *within* race groups. Unfortunately the data suggests that racial earning inequalities still remain a large problem.

Figure 6 shows lowess (locally weighted) regressions for mean hourly wages against highest level of education completed for males. It can be seen that white males earn more compared to their equally educated African counterparts. This could perhaps be because earnings inequality in South Africa is driven by educational quality to a greater extent than educational quantity differentials (Van der Berg, 2007: 1). In other words, South Africa's improvement in terms of quantitative educational attainment may not necessarily be conducive to earnings distribution equality (Lam, 1999: 11-12).

Figure 6: Lowess Regression on Mean Hourly Wages by Years of Education Completed, Males



3.2 Qualitative Educational Attainment

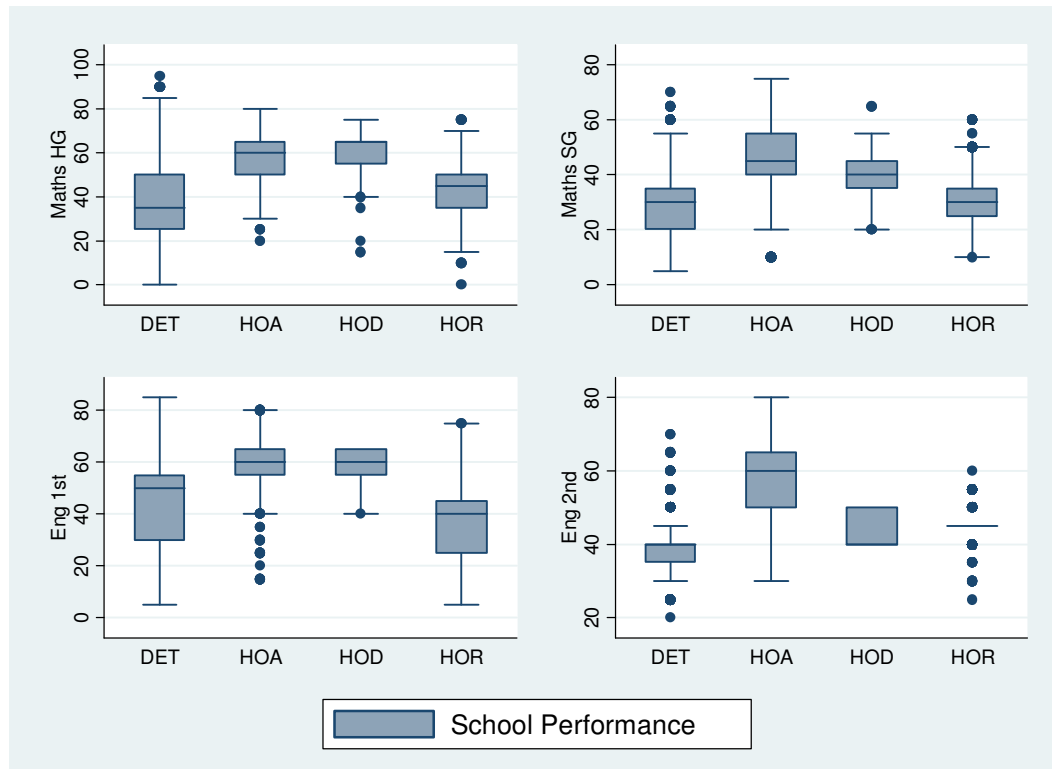
3.2.1 Analysing School Performance

Figure 7 shows box plots⁵ of school's performance on maths higher grade and standard grade and English first and second language by former racially-based departments. Under the apartheid regime the Department of Education and Training (DET) was responsible for African children; the House of Assembly (HOA) served white children; the House of Delegates (HOD) looked after Indian pupils; and the House of Representatives (HOR) was responsible for the coloured population group. Although

⁵ Box plots indicate a variable's locality and variation and consequently provide a simple way to compare the distribution of a variable for different groups that are plotted next to each other (Wood, 2001: 19; van der Berg *et al.*, 2002: 292).

there has been a large influx of African children into previously HOA schools in recent years, 96% of African children still attend previously DET schools. As already mentioned in section 2.1 the output-based approach to educational quality measurement is used in this section. Here school performance levels are used as proxy for quality of education.

Figure 7: Box Plots of School Performance by Former Racially-Based Departments

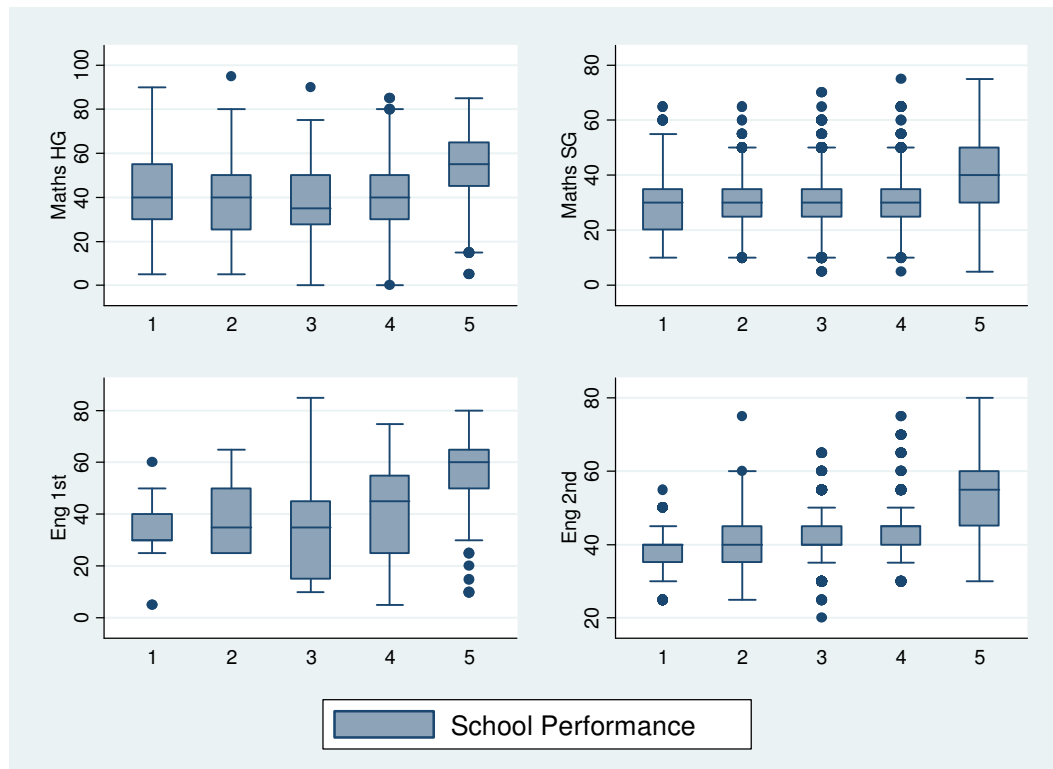


Since the sample size for previously Indian schools is relatively small, this section focuses mainly on the remaining three former departments. In the top left hand corner it can be seen that median maths higher grade scores – represented by the line within the box – were significantly better for previously white schools than for previously coloured and African schools. It is furthermore interesting to note that the bottom line of previously white schools coincides with the top lines of the boxes for previously African and coloured schools. This implies that 75 percent of previously white schools had an average maths higher grade mark above 50 percent, whereas 75 percent of historically African and coloured schools scored 50 percent or less on maths higher grade.

Schools’ performance on the remaining subjects display similar patterns. Previously white schools’ median scores on maths standard grade as well as English first and second language were far above those of historically coloured and African schools. The large vertical range of the box plots for previously African and coloured schools’ performance on English first language, especially in the lower tail, indicate greater variation.

The box plots in Figure 8 suggest that there still exist large qualitative differentials between the wealthiest and the poorest schools; the median scores for the richest schools (fifth quintile) is significantly better than the rest of the schools for all 4 subjects.

Figure 8: Box Plots of School Performance by School Quintile



It should furthermore be pointed out that these box plots display relatively poor performance on mathematics across all the schools. Since relatively high marks in mathematics are required in order to qualify for tertiary studies in, amongst other things engineering, science and commerce, it can be argued that the education system still largely fails to improve the skills constraint problem in South Africa and furthermore does not provide adequate opportunity for upward mobility of previously disadvantaged children in the labour market (Van der Berg, 2007: 9).

As mentioned in section 2.1, an overall performance score, based on a school’s performance on the above mentioned subjects, was assigned to each school in the dataset. The performance scores ranged between 0 and 0.45. Table 3 and 4 display the frequency distribution of schools’ performance by former racially-based department and school quintile. These tables basically display the same trends as shown by the previous box plots.

Table 3 confirms that historically disadvantaged schools lag considerably behind previously white schools in terms of overall performance. More than half of previously white schools scored 0.25 or

higher on overall performance, as against just over 2 percent of historically African schools and less than 1 percent of historically coloured schools.

Table 3: Frequency Distribution of Schools by Former Racially-Based Department and Performance

	DET		HOA		HOD		HOR		Total	
	number	%	Number	%	Number	%	number	%	Number	%
0	403	20.05	20	3.89			56	8.46	479	14.67
0.05	741	36.87	42	8.17	3	3.80	159	24.02	945	28.94
0.1	589	29.30	26	5.06	19	24.05	319	48.19	953	29.19
0.15	176	8.76	64	12.45	11	13.92	94	14.20	345	10.57
0.2	53	2.64	92	17.90	14	17.72	29	4.38	188	5.76
0.25	44	2.19	145	28.21	32	40.51			221	6.77
0.3	3	0.15	99	19.26			5	0.76	107	3.28
0.35	1	0.05	23	4.47					24	0.74
0.4			1	0.19					1	0.03
0.45			2	0.39					2	0.06
Total	2010	100	514	100	79	100	662	100	3265	100

Source: Own calculations from NIDS 2008

In Table 4 it can be seen that schools in the bottom 4 quintiles are clustered around the lower half of the performance distribution, this is in stark contrast with schools in the top quintile.

Table 4: Frequency Distribution of Schools by School Quintile and Performance

	1		2		3		4		5		Total	
	number	%	number	%	Number	%	number	%	number	%	number	%
0	194	19.84	142	18.91	200	16.84	88	10.96	55	8.46	679	15.54
0.05	394	40.29	267	35.55	398	33.50	277	34.50	63	9.69	1399	32.01
0.1	276	28.22	278	37.02	377	31.73	247	30.76	76	11.69	1254	28.70
0.15	104	10.63	58	7.72	142	11.95	99	12.33	60	9.23	463	10.59
0.2	10	1.02	5	0.67	56	4.71	42	5.23	93	14.31	206	4.71
0.25			1	0.13	12	1.01	49	6.10	177	27.23	239	5.47
0.3					3	0.25	1	0.12	99	15.23	103	2.36
0.35									24	3.69	24	0.55
0.4									1	0.15	1	0.02
0.45									2	0.31	2	0.05
Total	978	100	751	100	1188	100	803	100	650	100	4370	100

Source: Own calculations from NIDS 2008

Performance inequalities can largely be ascribed to variation in quality of education. These results consequently suggest that, despite large post-apartheid resource shifts, the South African educational system is still plagued by the persistence of former qualitative educational inequalities.

These findings are in accordance with findings by Van der Berg (2007: 14) who explored the matriculation pass rates of 1999 and 2000 and found that there are “massive differentials between the poorest and the richest school groups (average pass rates of 44% versus 97%), and predominantly Black - and predominantly white schools (43% versus 97%).” Since 80 percent of school-going children attend previously African schools, these schools’ performance are central to the improvement of education in South Africa (Van der Berg, 2007: 3-4).

3.2.2 Analysing Numeracy Test Scores

The analysis in the previous sub-section concentrated explicitly on a specific educational level, i.e. grade 12. The numeracy test scores provided by NIDS allow researchers to analyse cognitive skills across respondents that was exposed to different stages of South Africa’s complex educational history. In order to put the numeracy scores into perspective, it should be taken into consideration that the test scores range between -4.92 and 3.21 and the mean test score for everyone who took the test is -0.50.

Figure 9: Kernel Density of Numeracy Test Score by Race

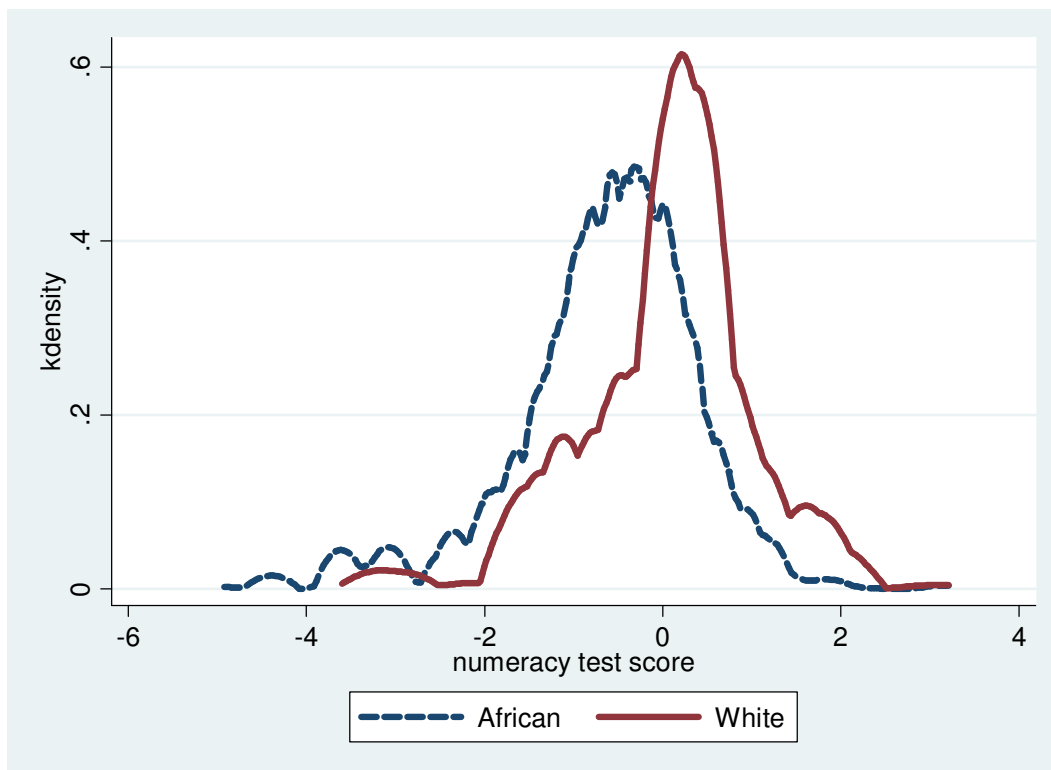


Figure 9 shows the kernel density distributions of numeracy test scores across the African and white population groups. These estimates are based on the entire sub-sample that was eligible to write the test and did indeed do so. It can be seen that there was more variation in the test scores for Africans since the curve does not peak as sharply as that of the whites. The distribution for whites furthermore lies to the right which implies that white individuals' performance on the numeracy test score exceeds those of their African counterparts.

Figure 10 shows lowest (locally weighted) regression curves for numeracy test scores against years of education completed for different race groups in South Africa. This figure allows one to determine the cognitive differential for individuals with the same level of education, from different race groups. It can be seen that Africans underperformed compared to their white counterparts. The racial differentials in terms of cognitive skill can mainly be ascribed to differences in the quality of education received by different race groups

Figure 10: Lowess Regression on Numeracy Test Score by Years of Education Completed and Race

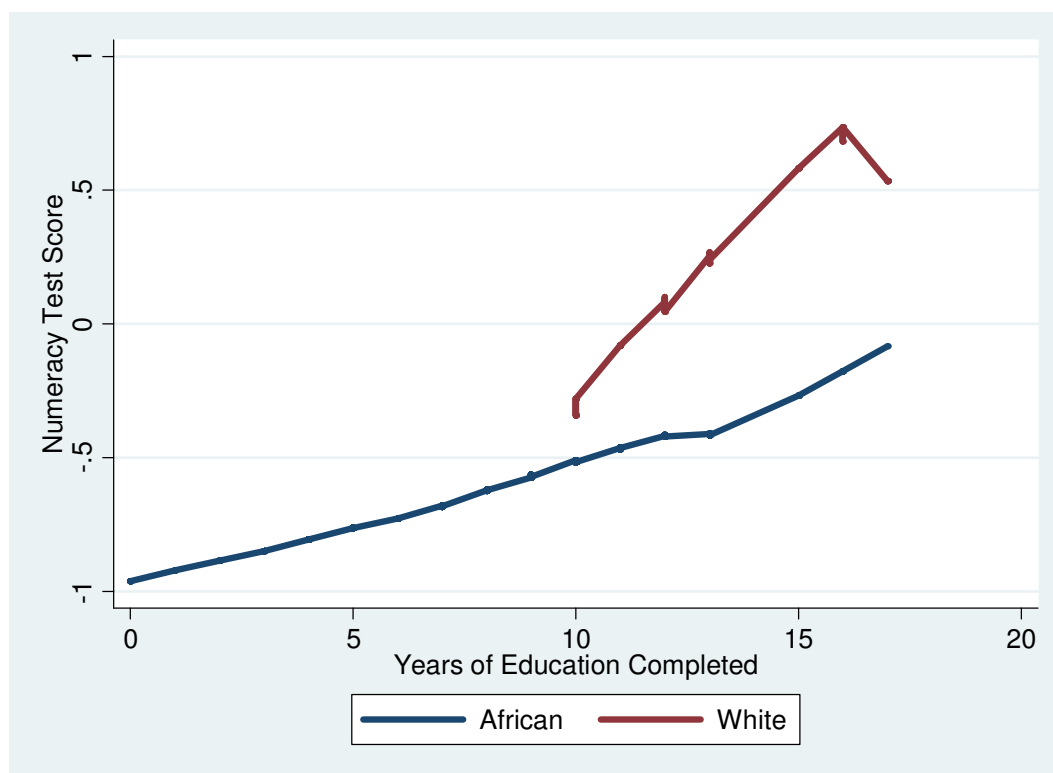


Table 5 disaggregates numeracy test scores by race and age cohort. As already mentioned, and as shown by the small sample sizes in Table 5, great care should be taken when interpreting these results.

Table 5: Mean Scores on Numeracy Test, by Age Cohort and Race

	African			Coloured			Indian			White		
	Number of obs	Mean	Std dev	Number of obs	Mean	Std dev	Number of obs	Mean	Std dev	Number of obs	Mean	Std dev
15-19 year olds	1135	-0.54	1.04	125	-0.23	0.97	5	0.29	0.63	10	-0.09	0.80
20-24 year olds	622	-0.60	0.90	86	-0.36	1.37	4	-0.18	0.49	11	-0.19	0.80
25-29 year olds	369	-0.69	1.00	78	-0.47	1.05	1	-3.08	.	7	-0.01	0.50
30-34 year olds	257	-0.50	0.82	59	-0.08	1.03	3	-0.19	0.49	10	-0.07	0.43
35-39 year olds	169	-0.68	1.09	63	-0.31	1.00	1	-0.18	.	17	-0.09	0.72
40-44 year olds	109	-0.68	1.17	43	-0.79	1.25	1	-0.79	.	11	0.23	0.48
45-49 year olds	98	-0.93	1.24	38	-0.12	0.75	3	-0.41	0.61	15	0.27	0.58
50-54 year olds	48	-0.85	1.00	31	-0.72	1.54	0			22	0.37	1.72
55-59 year olds	22	-1.03	1.21	11	-0.47	1.28	0			13	0.53	0.62

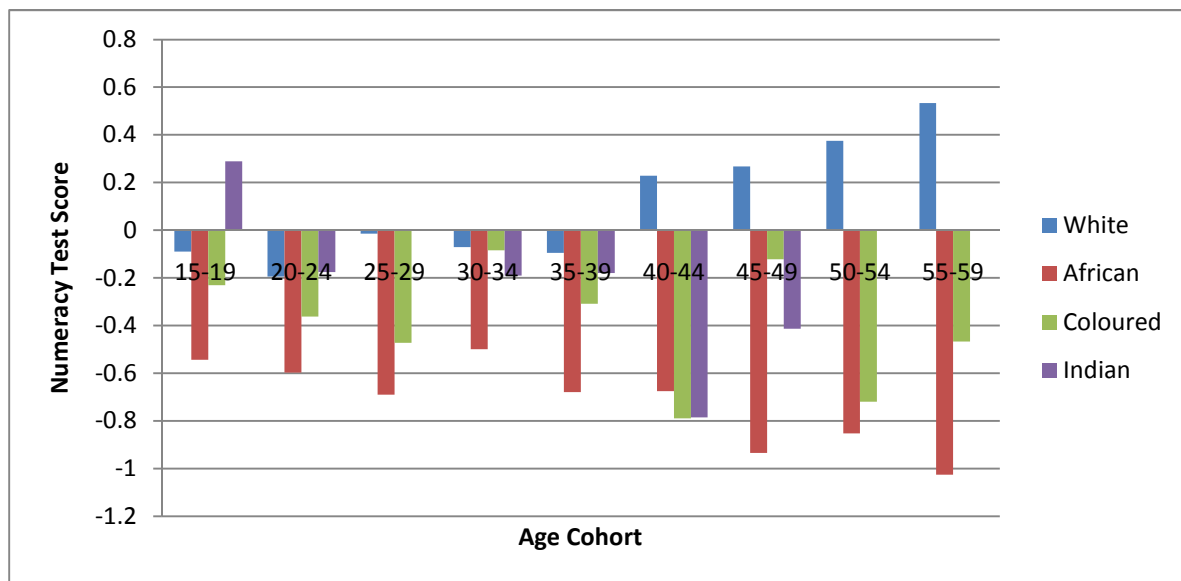
Source: Own calculations from NIDS 2008

Figure 11 graphically shows that cognitive differentials decrease as one moves towards the younger cohorts. The test score differential between Africans and whites amongst the 55-59 year old cohort is 1.56, compared to a differential for the 15-19 year old cohort, this drops to 0.45.

The low test scores for the older African cohorts are largely driven by the fact that the mean years of education for these cohorts are extremely low. As shown in Table 2, Africans in the 55 to 59 year old cohort had attained only 34.75% of the years of education reached by their white counterparts. Low test scores for the older African cohorts therefore comes as no surprise, since an individual with fewer years of education will perform worse on a numeracy test than one who had more years of schooling.

The same argument can however not be used for the younger African cohorts. This is because mean years of education completed for these younger cohorts is not significantly different to that of their white counterparts. It therefore seems that the test score differential between these race groups is driven by variation in school quality.

Figure 11: Mean Numeracy Test Scores, by Age Cohort and Race



Source: Own calculations from NIDS 2008

3.3.3 Regression Analysis

In order to control for a variety of factors that may have affected the outcomes of numeracy test scores, regression analysis were conducted. These were in the form of regressions with numeracy test score as dependent variable that was run by adding variables successively.

The first set of regressions was executed for the entire eligible sub-sample, i.e. respondents between the ages of 15 and 59; Table 12 in Appendix B displays the results. These results promote an understanding of the factors that have an effect on numeric ability, across a wide range of ages and among respondents that were educated during different stages of South African education (Griffin *et al.*, 2010: 1).

The second set of education production functions only considered those between the ages of 15 to 21 year olds; the upper-bound is an age at which most people would have completed their secondary schooling. Since school-going children's results represent the quality of education currently received, these regression are ultimately invaluable for the purposes of improving educational policies (Van Broekhuizen & Von Fintel, 2010: 4).

Finally, Heckman's two-step procedure is applied in order to control for possible sample selection bias. As discussed in section 2.2, there might be systematic differences between the subsample that took the test and the wider population. In order to draw conclusions about the wider population and

not just the subsample, special care should be taken to correct for possible selection bias (Kennedy, 2003: 284).

To correct for sample selection bias, Heckman proposed a two-step model in terms of which, applied to this particular case, the probability of completing the test is first modeled on the entire sample followed by the OLS model on numeracy test scores (Burger, 2008: 5; Kennedy, 2003: 284-285). A correction term, the inverse Mill's ratio (λ), is derived from the first step of the model. In order to circumvent the problem of sample selection, λ is included as an extra explanatory variable in the educational production function (Burger, 2008: 5; Kennedy, 2003: 824).

Regression 1 shows that years of education alone can explain 6.3% of the variation in numeracy test scores. An additional year of education raised an individual's numeracy score by 0.10. This coefficient was significant at all conventional levels.

To allow numeracy scores to vary by race, in the second regression, race dummy variables were added to the model. In column 2 it can be seen that the coefficient on education is smaller than in the first regression, which implies that some of the effects that were attributable to race were being captured by education. The positive coefficient on the white and coloured dummy variables implies that, given the same levels of education, a white or coloured individual performed better on the numeracy test than an African individual.

In the third regression, age was added to the model in order to allow for the fact that different age groups were educated during different stages of the South African educational system. The negative coefficient on age implies that older individuals performed worse on the numeracy test than younger cohorts, *ceteris paribus*. This may be because older cohorts received an education of lesser quality, but could also be because the cognitive skill tested – numeracy – tend to erode with age and less active use.

Regression 4 is a more reliable estimation of the racial gap in numeracy test scores since it controls for a wider variety of factors. It shows that numeracy test scores were affected by education, race, age, gender and geography. The signs on all the coefficients are in accordance with what one would expect.

Finally the estimation obtained by applying Heckman's two-step procedure is given in column 5 of Table 12 (Appendix B). The inverse mills ratio, λ is significant at a 10% level; this indicates that sample selection bias was in fact a relevant problem. This can be considered as being the most reliable estimation of the racial gap in numeracy test scores since it controls for selection bias.

In column 5, the estimated coefficient on the coloured dummy is slightly smaller than in column 4, which indicates that selection was biased upwards for the numeracy test performance of coloured individuals. The larger estimated coefficient on the white dummy indicates that selection was biased downwards for the performance of white individuals. Nevertheless, the coefficients on coloured and white dummy variables remained positive after controlling for selection bias. This implies that coloureds and whites performed better on the numeracy test than their African counterparts, *ceteris paribus*.

It should be noted that the insignificance of the coefficient on the Indian dummy variable, does not necessarily have any real economic meaning but is rather caused by the low concentration of Indian respondents in the sub-sample that took the test.

According to this estimation it is expected that an African male who has matric and lives in a rural formal area will score on average -0.02 on the numeracy test score, while his white counterpart is expected to score 0.71.

The same set of regressions was run for individuals between the ages of 15 and 21; Table 13 in Appendix B presents the results. Regressions 6 to 9 show that the results for the child population display roughly the same trends as those for entire population. As with the whole population, a Heckman two-step was applied to the most comprehensive regression, i.e. regression 9. Column 5 in Table 13 displays the estimation that was obtained after controlling for possible sample selection bias. The significance of λ indicates that the sample was actually biased.

The estimated coefficients indicate that, after controlling for selection bias, children's numeracy test scores were affected by years of education completed, race, age and gender. The positive coefficients on the coloured, Indian and white dummy variables imply that there still exist large racial differences in terms of educational quality. It should furthermore be pointed out that the coefficients on the coloured, Indian and white dummy variables were larger in regression 5 than in regression 4; this implies that selection was biased downwards for the numeracy test performance of these race groups.

When comparing the two results based on the Heckman selection model, it can be seen that children experience the racial effect more severely; the estimated coefficient on the white dummy variable is more than two as large for children as for the entire population. Education also seems to have a larger effect on children's outcomes (the coefficients on education are larger), which is in

accordance with what one would expect since more recent exposure to a school environment may give children an advantage.

4. Conclusion

An extensive empirical analysis of the NIDS dataset shows that the post-apartheid government has been largely successful in narrowing the quantitative educational gap caused by South Africa's divided past. Considering South Africa's history of segregation and discrimination, higher levels of educational attainment for previously disadvantaged individuals are most certainly a desirable outcome. Unfortunately, racial differentials on the numeracy test scores and matriculation results suggest that higher levels of educational attainment do not necessarily reflect positively on learner performance. It can consequently be concluded that, despite large post-Apartheid resource shifts, the South African educational system is still characterised by large qualitative educational differentials. This should be a cause of serious concern for policy makers, since poor quality of education adversely affects one's opportunities in the labour market and consequently hampers the upward mobility of previously disadvantaged individuals.

5. References

- Ardington, C. & Lam, D. 2007. *Suggestions for the Education Module of the South African National Income Dynamics Study*. [Online]. Available: www.nids.uct.ac.za [accessed 17 January 2011]
- Burger, C. 2008. *Sample selection bias and the South African wage function*. Stellenbosch Economic Working Paper 18/08. Department of Economics: University of Stellenbosch
- Chisholm, L. *The quality of primary education in South Africa*. Background Paper prepared for UNESCO Education for All Global Monitoring Report. 13th April 2004. [Online]. Available: <http://unesdoc.unesco.org/images/0014/001466/146636e.pdf> [accessed 1 May 2010]
- Griffin, P., Leibbrandt, M., Pavlovic, M., & Zuze, T. L. 2010. *Numeric Literacy in South Africa: Report on the NIDS Wave 1 Numeracy Test*. Technical Paper no. 5. [Online]. Available: www.nids.uct.ac.za [accessed September 2010]
- Gustafsson, M. *Using the hierarchical linear model to understand school production in South Africa*. Stellenbosch Economic Working Paper 01/07. Department of Economics: University of Stellenbosch
- Hanushek, E.A. 2007. *Education Production Functions*. Hoover Institution: Stanford University
- Hanushek, E.A & Woessmann, L. 2008. The Role of Cognitive Skills in Economic Development. *Journal of Economic Literature*, 46:3, 607-668
- Hedges, L. V. & Nowell, A. *Changes in the Black-White Gap in Achievement Test Scores*. *Sociology of Education* 72(2)
- Human Science Research Council. *South African component of TIMSS 2003*. [Online]. Available: http://www.hsrc.ac.za/Research_Programme-Page-61.phtml [accessed 22 January 2011]
- Kennedy, P. *A Guide to Econometrics*. Blackwell Publishing: Australia
- Keswell, M. & Poswell, L. 2004. *Returns to education in South Africa: A retrospective sensitivity analysis of the available evidence*. *South African Journal of Economics* 72(4)
- Lam, D. 1999. *Generating Extreme Inequality: Schooling, Earnings, and Intergenerational Transmission of Human Capital in South Africa and Brazil*. Research Report 99-439, Ann Arbor: Population Studies Center, University of Michigan.

Leibbrandt, M., Woolard, I., de Villiers, L. 2009. *Methodology: Report on NIDS Wave 1*. Technical Paper no 1. South African Labour & Development Unit. School of Economics: University of Cape Town. [Online]. Available: www.nids.uct.ac.za [accessed May 2010]

Louw, M., S. Van der Berg and D. Yu. 2005. *Educational Attainment and Intergenerational Social Mobility in SA*. Stellenbosch Economic Working Paper 09/06. Department of Economics: University of Stellenbosch

Taylor, S. & Yu, D. *The importance of socio-economic status in determining educational achievement in South Africa*. Stellenbosch Economic Working Papers 01/09. Department of Economics: University of Stellenbosch.

Van der Berg, S. 2007. Apartheid's Enduring Legacy: Inequalities in Education. *Journal of African Economies*, 16(5): 849-880, 2 August 2007

Wood, L. 2001. *A Statistical Analysis of Differentiation in the Education of South African Youth Based on the Living Standards and Development Survey 1993*. Stellenbosch. University of Stellenbosch. Assignment presented in partial fulfilment of the requirements for the degree of Masters of Commerce

Van der Berg, S., Wood, L., & le Roux, N. 2002. Differentiation in black education. *Development Southern Africa* 19(2): 289-306, June 2002

Van der Berg, S. 2006. *How Effective are Poor Schools? Poverty and Educational Outcomes in South Africa*. Paper to SACMEQ International Invitational Conference. September 2005. International Institute for Educational Planning, Unesco, Paris. Working Paper 6/2006. Bureau for Economic Research/Department of Economics: Stellenbosch.

Van der Berg, S & Louw, M. 2007. *Lessons learnt from SACMEQII: South African student performance in regional context*. Working Paper 16/07. Bureau for Economic Research/Department of Economics: Stellenbosch.

Van Broekhuizen, H & Von Fintel, D. 2010. *Who Responds to Voluntary Cognitive Tests in Household Surveys? The Role of Labour Market Status, Respondent Confidence, Motivation and Culture of Learning in South Africa*. Stellenbosch Economic Working Papers 27/10. Department of Economics: University of Stellenbosch.

Appendix A:

Table 7: Proportion Completing Grade 7

	African	White	Total
15-19 year olds	91.5	97.86	92.3
20-24 year olds	93.72	99.12	94.2
25-29 year olds	91.16	100	91.98
30-34 year olds	85.97	100	87.35
35-39 year olds	78.44	100	81.91
40-44 year olds	68.07	100	74.12
45-49 year olds	54.97	100	64.38
50-54 year olds	49.04	100	60.05
55-59 year olds	35.8	100	52.29
60-64 year olds	32.59	100	46.84
65-69 year olds	21.85	100	43.32
70-74 year olds	18.51	99.93	44.1
75-79 year olds	6.73	98.03	37.82

Source: Own calculations from NIDS 2008

Table 8: Proportion Completing Grade 12

	African	White	Total
20-24 year olds	39.61	68.51	42.72
25-29 year olds	39.62	73.76	43.1
30-34 year olds	39.45	75.3	43.73
35-39 year olds	31.95	82.02	39.11
40-44 year olds	22.91	80.6	31.55
45-49 year olds	15.43	87.8	28.71
50-54 year olds	14.64	83.57	27.27
55-59 year olds	3.71	74.44	18.39
60-64 year olds	5.03	56.43	15.77
65-69 year olds	3.26	59.05	17.61
70-74 year olds	4.16	56.28	18.03
75-79 year olds	0.64	46.79	13.44

Source: Own calculations from NIDS 2008

Appendix B

Table 12: Regressions of determinants of numeracy test score, Whole Population

	OLS 1	OLS 2	OLS 3	OLS 4	Heckman 1
Numeracy					
Education	0.1033***	0.0904***	0.0900***	0.0904***	0.0587***
Coloured		0.2601***	0.2799***	0.2668***	0.2645***
Indian		-0.05	-0.01	-0.03	0.51
White		0.5116***	0.5742***	0.5472***	0.7239***
Age			-0.0042**	-0.0029*	0.01
Female				-0.0768**	-0.1382***
Tribal Authority Areas				0.13	0.1466**
Urban Formal				-0.02	0.00
Urban Informal				-0.3147***	-0.13
Constant	-1.6100***	-1.5252***	-1.4120***	-1.3856***	-0.7218*
Select					
Education					0.0782***
Coloured					0.0708**
Indian					-0.8866***
White					-0.4175***
Age					-0.0270***
Female					0.03
Children					-0.0182**
Very Good Health					-0.0518*
Good Health					0.03
Fair Health					-0.03
Poor Health					-0.3620***
Employed					-0.1103***
Household Income					0.00
Constant					-0.5111***
Mills					
Lambda					-0.5240*
N	3489	3488	3488	3488	14497
F-stat	233.74	74.89	61.25	43.13	
Prob > F	0.00	0.00	0.00	0.00	0.00
R-squared	0.06	0.08	0.08	0.10	

Note: ***p<0.01, **p<0.05, *p<0.1

Own calculations from NIDS 2008

Table 13: Regressions of determinants of numeracy test score, Child Population

	OLS 6	OLS 7	OLS 8	OLS 9	Heckman 2
Numeracy					
Education	0.0861***	0.0869***	0.1179***	0.1268***	0.0750**
Coloured		0.3712***	0.3771***	0.3979***	0.5499***
Indian		0.6244**	0.7143***	0.7358***	1.3784**
White		0.1839	0.1692	0.1628	1.4799**
Age			-0.0559***	-0.0528***	-0.0509***
Female				-0.0161	-0.1647**
Tribal Authority Areas				0.1329	0.1757
Urban Formal				-0.0205	0.0448
Urban Informal				-0.3548***	-0.1305
Constant	-1.3817***	-1.4182***	-0.7250***	-0.8516***	1.0401
Select					
Education					0.0638***
Coloured					-0.2445***
Indian					-0.8494***
White					-1.1261***
Female					0.0091
Seldom Hungry					0.1343*
Sometimes Hungry					0.0753
Often Hungry					0.0272
Always Hungry					0.257
Household Income					0
Very Good Health					-0.1088**
Good Health					0.0142
Fair Health					-0.0558
Poor Health					-0.6149***
Constant					-0.7710***
mills					
lambda					-1.5414***
N	1613	1613	1613	1613	3560
F-stat	37.88439	13.81962	13.91656	12.689	
Prob > F	0	0	0	0	0
R-squared	0.02298	0.03323	0.0415	0.0665	

Note: ***p<0.01, **p<0.05, *p<0.1

Own calculations from NIDS 2008