The effect of neighbourhoods and school quality on education and labour market outcomes in South Africa

ASMUS ZOCH

Stellenbosch Economic Working Papers: WP08/2017

www.ekon.sun.ac.za/wpapers/2017/wp082017

September 2017

KEYWORDS: South Africa, Education, Spatial analysis, Neighbourhood effects, Family effects
JEL: D10, I24, I26, J15

ReSEP (Research on Socio-Economic Policy)
http://resep.sun.ac.za

DEPARTMENT OF ECONOMICS
UNIVERSITY OF STELLENBOSCH

SOUTH AFRICA

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

www.ekon.sun.ac.za/wpapers

The effect of neighbourhoods and school quality on education and labour market outcomes in South Africa[1]

Asmus Zoch[2]

ABSTRACT

This study evaluates the relative importance of family, neighbourhood and school quality in explaining variation in standardized test results, reaching and passing matric, university attendance and labour market earnings. It adds to the literature, by using a spatial approach to link a neighbourhood wealth index from the Census 2011 community survey to a unique administrative school data set from the Western Cape. For the long-term perspective the household and school information from the National Income Dynamics Study are explored. The results from administrative school data show how student wealth and differences in school quality produce vastly different outcomes for a cohort of grade 6 to 12 learners in Cape Town. It shows how grade 6 children going to the richest 20% of all schools are 30% more likely to pass matric in time, furthermore by grade 9 the learning gap is approximately four grade-levels worth of learning in comparison to children going to the poorest 20% of schools. However, this study also demonstrates that even children from the poorest neighbourhood would perform well if they go to one of the richest 20% of schools. Yet, given the limited number of quality schools, the segregated location of quality schools, financial as well as transport constraints, only very few children from the poorest 60% actually attend a top quintile schools. These results can be replicated for the national data set and show that in order to achieve more equal education outcomes, the quality of schools in the poor neighbourhoods have to be drastically improved. In addition, using the new school wealth index as an instrument for school quality, there seems to be a significant premium for quality education in labour markets earnings regressions, which show the long-term implications of the schooling system.

[1] The author is grateful for comments from Servaas van der Berg (SU) and Rulof Burger (SU) on earlier versions of the paper. The paper has been partly funded and written at the University of Stellenbosch and with support of ReSEP.
[2] zoch@uni-mannheim.de; Department of Economics, Mannheim University, L 7, 3-5, Room 123.

1 Introduction

The extent to which family, neighbourhood and the quality of schooling can explain differences in socio-economic outcomes is an important question that has produced several studies in the past (e.g. Card and Krueger, 1992; Case and Deaton, 1999; Case and Yogo, 1999; Solon et al., 2000; and Altonji and Mansfield, 2011). In light of recent findings by Lam et al. (2011) that movement through secondary schooling is still predominantly along racial lines, the effect of special location and schooling outcomes is a particularly important research question for South Africa. In addition, South Africa provides an interesting setting to observe neighbourhood and school effects, due to its unique history and sharp differences between wealth and school quality in different neighbourhoods, which were shaped under the apartheid regime (Case and Yogo, 1999). Given South Africa's high income inequality levels, making sure that child from poor socio-economic backgrounds and former disadvantaged areas also receive quality education is essential to transform the education system into a resource for increased opportunity and racial equity.

Because of the strong correlation between family wealth, school choice, neighbourhood and schooling outcomes, it is very difficult to disentangle the relative importance of each factor. Previous studies analysing school and neighbourhood effects did not have the necessary data to observe school quality, neighbourhood wealth and household characteristics at the same time. Furthermore, since these factors are highly correlated, imperfect measurement of one will make the other factors seem more important than they really are. The only way in which the relative importance of these education and wage determinants can be accurately measured is to find more reliable measures of each of these factors.

This study contributes to the literature by building a model which explains schooling and labour market outcomes, given more informative measures of students' location, parents' background as well as school quality measures. To do so, a unique administrative school data set, the Centralized Education Management Information System (CEMIS), is used to follow a cohort of Grade 6 learners in Cape Town. The CEMIS data as well as the National Income Dynamics Study are then also linked to the master list of schools in South Africa, as well as to Census 2011 community data. Spatial linking of the Census 2011 allow us to construct a new wealth index for 85 000 small areas and identify precisely the neighbourhood wealth of each household and school in the data. It will be shown that this wealth measure for each school is an excellent proxy for school quality, which explains schooling outcomes more accurately than the official school quintiles provided by the department of education (this is true at least in metropolitan regions). Finally, the data allows us to observe whether a child is going to school in the same neighbourhood he/she is living in and the actual distance he/she travels to school every day.

Having constructed the data in this unique way, this paper will try to answer some questions that have important policy implications. How large are family background effects relative to school quality effects when explaining schooling outcomes? What is the effect of sending a child from the poorest

neighbourhoods to the best schools in the region? Are there long-run effects of going to a high quality school that determine university enrolment and an earnings premium? While this paper will show large differences in education outcomes depending on the quality of the school, it can also demonstrate that even children from the poorest neighbourhood would perform well if they went to these high performing schools. However, given the limited number of quality schools in the country and the financial as well as transport constraints faced by the poor, only about 3-10% of the poorest 60% of children actually manage to get high quality education. Hence, to achieve more equal education outcomes, the quality of schools in the poor neighbourhoods would need to be improved. This would have large effects, as it can be shown that there are indirect effects as well as a direct premium for quality education in the labour market.

The structure of this paper will be as follows: First, a short background and literature review is given. Second, the four main data sets and the merging process with which the datasets were linked will be explained, and descriptive statistics and maps provided. Lastly, a regression analysis for education outcomes and labour market earnings is performed. Finally, a conclusion about the findings is drawn.

2 Background and Literature

One of apartheid's enduring legacies is the large inequality of education opportunity for children from different racial and socio-economic backgrounds. This situation is rooted in the apartheid school system that created separate departments for children of different race groups: white, Indian, coloured and black (Hill, 2016). The result was four school systems within South Africa with vastly different resources[3], curricula and learning environments. While transforming the education system to achieve equal opportunity has been an important policy of the post-apartheid agenda, the institutional memory of the former school departments is still causing significant differences between schools along racial lines (e.g. Van der Berg, 2007; Van der Berg et al., 2011; Taylor et al., 2013; and Yamauchi, 2011). In particular former black schools (in urban areas also called township schools) have not seen much racial mixing and are still under-resourced with numerous administrative problems. In the worst case, the situation at these schools can be described as a culture of learning where teaching is almost non-existent (Msila, 2005). By contrast, former white and Indian schools are now much more racially diverse, although not socio-economically diverse, due to their fee-charging structure (Yamauchi, 2011).

Today, because of the *exit* option and due to new laws stating that no child can be excluded from a school for financial reasons (Hunter, 2015), in theory all parents are allowed to send their children to former Model C schools (formerly white schools). However, during apartheid, different population groups were also segregated in separate residential areas which means good schools are still located in selected areas (Yamauchi, 2011). As a results, de Kadt et al. (2014) finds for the "Birth to Twenty cohort study" of 1428

[3] The situation in the 1960s was that black students received only one-tenth of the school funding as white children (Hunter, 2015).

children in Soweto-Johannesburg that over a third of them travel more than 6km to school, 60% leave the suburb they live in and only 18% attend their nearest school. However, household-level financial constraints caused by schooling fees, additional transport and time costs will in practice prevent the poorest children from attending better schools. A qualitative study by Msila (2005) interviewed parents currently sending their children to historically black schools. She observes that distance and economic reasons mostly prevent parents from sending their children to former Model C schools. In terms of racial patterns, Hill (2015) finds that coloured families in Cape Town are 50% more likely to send their children to "better" schools than black families.

Because of the strong correlation between family wealth, school choice, neighbourhood and peer effects, it is very difficult to disentangle the relative importance of each factor in schooling and labour market outcomes. That is, do schools produce different outcomes because they influence student performance or because they were able (or failed) to attract students that would have succeeded regardless of the school chosen? Similarly, do children in rich neighbourhoods perform well at school because of the neighbourhood they live in, the school they go to or the family they come from? The answer to this question is of particular importance due to its policy implications (Altonji and Mansfield, 2011). In the case where most of the variation of learning outcomes can be explained by parents' socio-economic status (SES) and their involvement in students' learning behaviour, increasing school funding of the poorest schools will be unlikely to improve student outcomes. On the other hand, if parents' SES is only significant because it is a proxy for the quality of the school their children go to, then improving school quality is likely to have large effects, in particular for the poorest students (Altonji and Mansfield, 2011).

To solve the correlation and identification problem, the seminal study by Solon et al. (2000) used a variance decomposition approach to bound the proportion of socio-economic outcomes that can be attributed to disparities in family and neighbourhood background by using the correlation between siblings and unrelated neighbours. The neighbourhood correlation captures the pure neighbour effect but also family traits because of the sorting mechanism and therefore is an upper bound[4]. Previous studies adapting the Solon et al. (2000) methodology found relatively high sibling correlations in brother income and education for China, the US, UK and Germany and smaller effects for the Nordic European countries (e.g. Björklund et al., 2004; Raaum et al., 2006; Lindahl, 2011; Eriksson and Zhang, 2012; Nicoletti and Rabe, 2013; or Schnitzlein, 2014). The proportion that can be explained by disparities in neighbourhood background seems to be nearly zero for income and small for education. Other studies for the US that focused on school quality and its importance in reducing disadvantage from one generation to the next, also found very small effects (e.g. Jenning et al., 2015).

[4] The sorting mechanism means that parents with similar characteristics will move into the same neighbourhoods. Therefore, there is a strong correlation between parents' characteristics and neighbourhood wealth.

This literature is in sharp contrast to studies by Card and Krueger (1992), Case and Deaton (1999), and Case and Yogo (1999) who find for the USA as well as South Africa that school quality measured by pupil/teacher ratios have large and significant effects on the return to schooling for black men. Hanushek, Lavy, and Hitomi (2008) show that children in Egypt were more likely to drop out of low quality primary schools, which is similar to findings by Harbison and Hanushek (1992) who observe a reduction in repetition rates for Brazilian primary students when going to high quality schools. Another study by Glewwe et al. (2014) show that whether or not there are differences in learning between children from different backgrounds at the same school depends on the type of students and the country context.

There are at least two possible reasons for the contrasting results in the literature on the role of school quality. First, while the first studies following the approach of Solon et al. (2000) explicitly tried to disentangle the effect of neighbourhoods, family and school quality by using sibling correlations, the later studies might not have efficiently separated family and school quality effects. Hence, those studies finding large schooling effects might be biased due to unobserved heterogeneity due to family effects. The other possibility is that as Case and Yogo (1999) rightly argue: "schooling quality should matter in countries like South Africa, where resources were distributed very unevenly between regions in the past" but these difference are not large enough in Western societies to matter much. In the specific case of South Africa where we do know do that there are basically two school systems – those of the former white and those of the former back schools – we should definitely observe some quality differences in schools. Given the close proximity of poor and rich neighbourhoods and the clear distinction between poor and rich schools and children travelling in between these boundaries, this study should shed light on the role of parental background, neighbourhood, and school quality. Furthermore, due to data constraints most studies, except Altonji and Mansfield (2011) or Jenning et al. (2015), focused on short run effects and did not observe the long-term impact of school quality, family and neighbourhood background. Again, having long-run panel data this study will be able to address this issue.

3 Data and descriptive statistics

To analyse the effect of neighbourhood and quality schooling two main data sets are used: the Western Cape administrative school data, CEMIS, and the National Income Dynamics Study (NIDS). In this part, it will be explained how CEMIS and NIDS are merged with Census 2011 community spatial data and the master list of schools in South Africa. Some descriptive information and maps of the merged data will be provided.

3.1 Census 2011

The Census 2011 data is provided by Statistics South Africa (StatsSA). The primary sampling units (PSUs) were the 103 576 Census enumeration areas (EA) (see StatsSA, 2012b). To obtain detailed information on the neighbourhood level the analysis was based on the "Small area layer" (SAL) from the Census 2011

Community Profiles which StatsSA provides in SuperCROSS. SAL are the second lowest geographical areas (after the EA level) in which the country is divided for the Census design. In a second step the information from the SALs were aggregated to weighted averages (using population size) to get the "Subplace" (SP) information. StatsSA (2012a) defines SP as "the second (lowest) level of the place name category, namely a suburb, section or zone of an (apartheid) township, smallholdings, village, sub village, ward or informal settlement". The Census community data provides information for about 85 000 SALs and 22 000 SPs in South Africa. However, our analysis showed that there are SALs which do not have sufficient household information given their population size.[5] For those 629 SALs no wealth index was constructed but rather the weighted averages from the SP were given. Using SAL level information therefore has the advantage of identifying data irregularities we would not observe on the aggregate SP level and having smaller and more even distributed area samples to construct a wealth index using principal component analysis (PCA).

An example of SP and SAL maps can be found in the Appendix (Figure A1). For dense areas like the City of Cape Town each SP is divided into about 16 new SALs. Between 0 and 111 937 individuals live in a SP. However, the median size is about 5 400 individuals living in 1 500 households for SP and about 580 individuals living in 160 households in a SAL.

One of the largest challenges in this approach comes from the fact that the data is provided on community level rather than household level. Therefore, all variables in SuperCROSS are given as total counts (e.g. the number of households with access to clean water in that particular SAL). These totals were transformed to percentages, e.g. the percentage of households with access to clean water in that SAL. The wealth index was then constructed using PCA analysis for a set of variables, namely household income (in income brackets), education (for everyone age 25-64), labour market status, household assets and household services. The household income is given in 12 brackets ranging from "R1-4800" to "R2 457 601 or more" and "no income". The four lowest income categories and the highest five categories were grouped and aggregated together. Similarly education was grouped as "No education", "Some primary education", "Secondary education", "Matric" and "Higher education". In addition only education levels for the age group 25-64 were chosen to observe completed education and not enrolment. For a complete list of all variables used to construct the wealth index, see Table A1 in the Appendix.

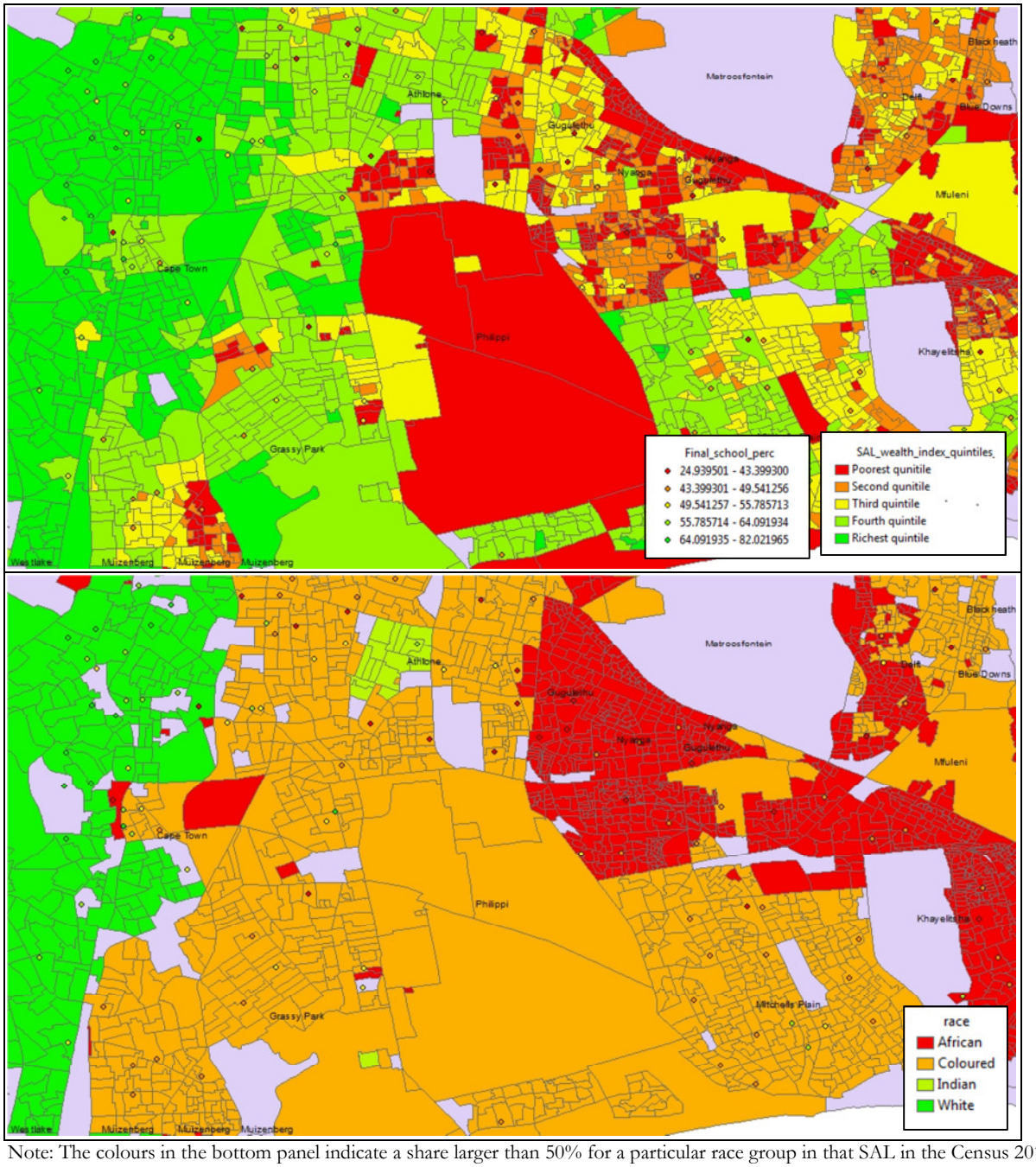
Figure A2 in the Appendix shows the distribution of the wealth index. To make the wealth index comparable to the official school quintiles provided by the department of basic education (DBE), quintiles of wealth were generated using the population size of each SAL and SP respectively as weights. Wealth quintile information has been merged with the national school master list and NIDS learner data using GPS coordinates and GIS software as described below.

[5] For more detailed information on the net Census coverage error see (STATSSA, 2012b).

3.2 Master list of schools

The master list of schools in South Africa provided by the DBE has detailed information for all 25 827 schools of the country, including ex-department, quintile, learner-teacher ratios and GIS information which are used for the matching process[6]. The upper panels of Figure 1 and Figure 2 show the newly calculated neighbourhood wealth quintiles and the average matric examination results for 2014 for the municipalities of Cape Town and Johannesburg.[7]

Figure 1: Neighbourhood wealth quintiles, geographic race distribution and matric 2014 examination results in metropolitan Cape Town (Source: Census 2011 and DBE school data)

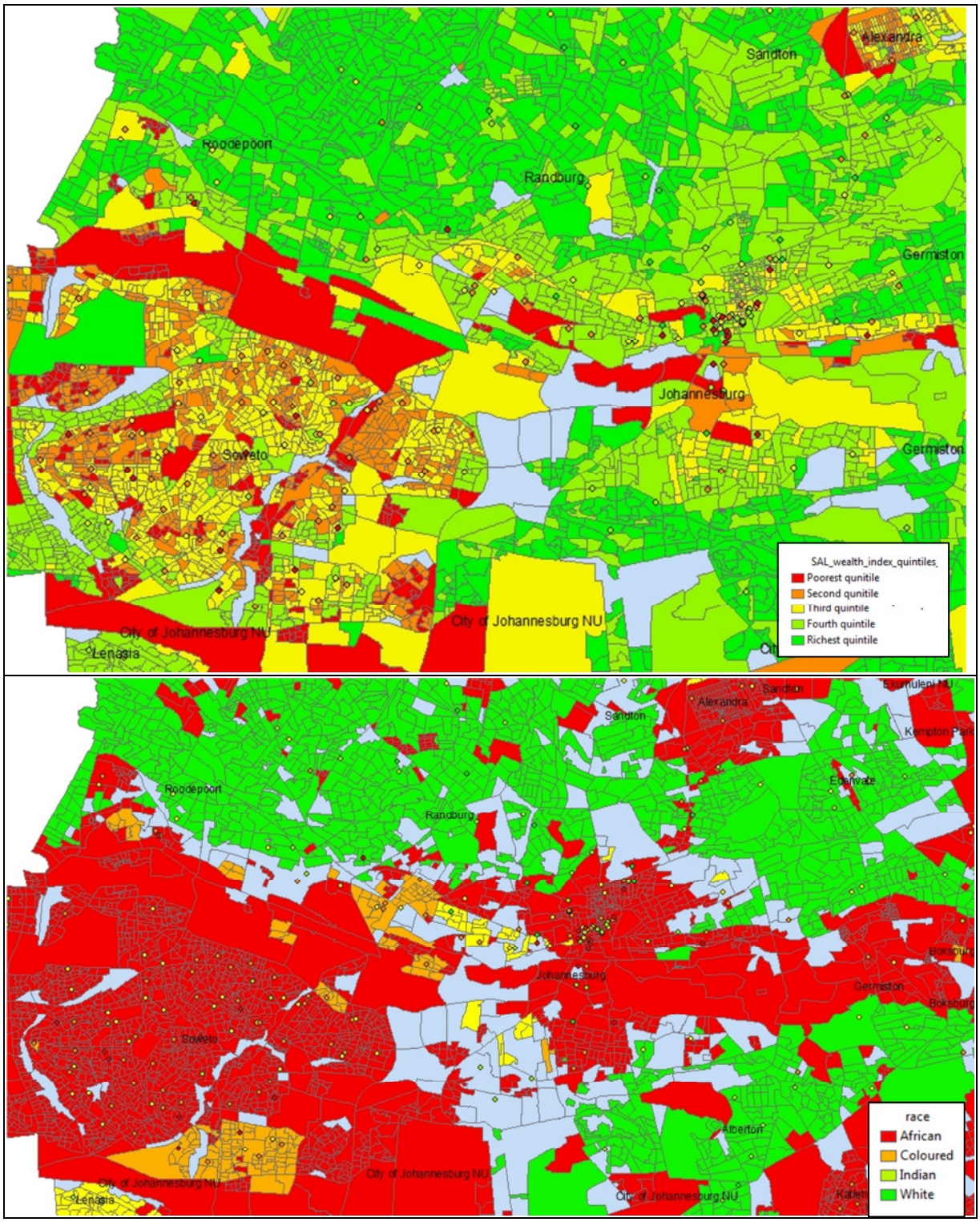


Note: The colours in the bottom panel indicate a share larger than 50% for a particular race group in that SAL in the Census 2011 data. In addition, the average matric 2014 examination results per school are displayed.

[6] For more information on the master list of schools in South Africa see Van Wyk (2015).

[7] For the metropolitan region of eThekwini see Figure C2 in the Appendix.

Figure 2: Neighbourhood wealth quintiles, geographic race distribution and matric 2014 examination results in metropolitan Johannesburg



Note: The colours in the bottom panel indicate a share larger than 50% for a particular race group in that SAL in the Census 2011 data. In addition, the average matric 2014 examination results per school are displayed.
(Source: Census 2011 and DBE school data)

The neighbourhood wealth quintiles are illustrated by colours: green for the richest and red for the poorest areas. Similarly, the matric examination results for 2014 are also divided into quintiles and coloured in green for best and red for worst performing schools. As mentioned before, due to the legacy

of apartheid there is a strong correlation between the neighbourhood a school is based in and the average school results it produces, which clearly can be seen in the maps.

In addition, the bottom panels of Figure 1 and Figure 2 illustrate how racially segregated South Africa still is today. That is, there appears to be some clustering for each race group where their share of the SAL population exceeds more than 50%. Hence, given the institutional memory of the schooling departments from the apartheid regime and the clear racial separation, we would expect some form of omitted variable bias in any schooling model which does not sufficiently control for household and schooling location.

3.3 CEMIS data

The Centralized Education Management Information System (CEMIS) is a unique administrative school date set, which was developed by the Western Cape Education Department to register and track learners through their school career (Van Wyk 2015). It is possible to get exact household location of nearly every learner, transforming street addresses into GPS coordinates to use them for Geographic Information System (GIS) analysis[8]. To simplify the interpretation of the results and to get rid of specific town and rural area effects, the analysis is limited to students attending a school located in the municipality of Cape Town, which leaves about two-third of the original provincial sample. Second, to observe learners for the longest period possible, the learner sample only includes the cohort of grade 6 in 2007 that can be observed until they reach grade 12 in 2013. For this cohort one can match their Western Cape Systemic Tests results of grade 6 and grade 9 [9] as well as their final matric results in 2013. In addition, CEMIS provides information on grade outcomes (failed, repeated, drop out) for each year which can be used to model learner success. In total, this leaves a Cape Town CEMIS sample of 45 567 students living in 3 605 different neighbourhoods and attending 710 different schools (for the cohort of grade 6 in 2007). A map of the learner location distribution can be found in Figure A4 and Figure A5 in the Appendix. It demonstrates that learners from all areas are included in the learner sample.

Table 1 provides the racial distribution and panel survival rates for the 2007 Cape Town grade 6 learner cohort. First, it shows that the student distribution is significantly different from the racial distribution of the country. The share of blacks (34% vs 38%) and whites (8% vs 11%) is slightly lower than for that age group in the 2011 Census data (see Table A2 in the Appendix).[10] Second, the survival rates to 2013 are low with an overall average of 57% of the original 2007 cohort remaining in the school system in 2013.

[8] To get the GIS information, the learner addresses were matched with official street records that had GPS coordinates for every street in Cape Town. Whenever the match was not successful, midpoints of the neighbourhood of the student were used instead. This should still provide accurate wealth estimates for each student but biases the distance to school estimates which will be taken into account in the regression analysis. Note that at all times student's identities and information were kept completely confidential and no individual child can be identified in the data set after the merges have taken place.

[9] For more information on Western Cape Systemic Tests see Van Wyk (2015).

[10] The reason for the under capturing might be particularly bad address information for some of the poorest black students, while some of the richest white students go to private schools which are not part of the CEMIS project. Dropping some of the richest and poorest students may lead to an underestimate of the true effects of neighbourhood wealth.

However, this is very much in line with estimates from the General Household Surveys (GHS) that about 43% of South African students entering grade 8 drop out of school before reaching matric (Taylor et al. 2011). Black children are even less likely to be still observed in 2013, with only 51% being observed again in the school system in 2013. For the coloured (61%), the Indian and Asian (66%) and in particular the white students (66%) the panel survival rates are significantly higher but overall still surprisingly low given their much better socio-economic status. Yet Table 2 shows that the main reason students leave the school system after grade 6 is not because they failed their grades. Therefore, one can assume that in particular for the richer (white) students most of drop outs are learners moving from a public primary school to a private secondary school between the years 2007/2009.

Table 1: Racial distribution and survival rates of the 2007 Cape Town learner distribution (Source: CEMIS)

Year	2007	2008	2009	2010	2011	2012	2013	Survival rate
Black	15235	13791	11174	10471	9507	8625	7701	51%
Coloured	24825	24101	22012	21203	19733	17692	15146	61%
Indian/Asian	466	424	354	354	328	317	307	66%
Other	1001	919	603	431	392	251	168	17%
White	3751	3551	2842	2801	2647	2538	2469	66%
Total number	45278	42786	36985	35260	32607	29423	25791	57%

On the other hand, for black students grade failure starts to become an important explanatory variable for leaving the school system in Grade 10 and Grade 11. A more advanced model of school drop outs and school success will be presented in the next section.

Table 2: Grade outcome the year the student was last observed in the school system (Source: CEMIS)

		2007/ 2008	2008/ 2009	2009/ 2010	2010/ 2011	2011/ 2012	2012/ 2013
No info	**Black**	29%	0%	0%	1%	1%	2%
	White	13%	0%	0%	2%	3%	3%
Passed grade	**Black**	65%	94%	89%	81%	53%	31%
	White	84%	99%	96%	90%	78%	79%
De-register	**Black**	1%	1%	4%	5%	4%	7%
	White	0%	0%	1%	1%	1%	2%
Failed grade	**Black**	5%	5%	7%	14%	42%	61%
	White	3%	2%	3%	8%	18%	15%

3.4 NIDS

Besides the publicly available data, NIDS also provides secured data on GPS household location, as well as the names of all the schools children are going to or the last school a respondent went to before leaving the school system. Using the name of the school, as well as the household location, it was possible to merge in further information (containing GPS data for each school) from the master list of schools using

fuzzy matching.[11] Having GPS coordinates for all households as well as schools makes it possible to link the new neighbourhood wealth index from the Census 2011 community data to the household and school location using the SAL and SP maps described in the last section. This data also allows us to calculate the distance to the closest school, as well as determining the actual school students attended.[12] While the CEMIS data allows tracking a larger cohort of students and matching student records to standardized test results, NIDS has the advantage of being representative at a national level, which allows validating the findings from Cape Town to the rest of the country. This is important, given the very unique racial composition of the Western Cape.

3.5 Learner wealth distribution

As mentioned before, due to the legacy of apartheid there is a strong correlation between the neighbourhood a school is based in and the average school results it produces. However, as Yamauchi (2011) has shown, there is also a strong positive correlation between school quality and student fees. Therefore, even if two schools are both based in the richest neighbourhood, their quality still might be significantly different if the one only has rich students that can pay high fees and finance extra teachers, while the other is mostly state funded and has a much higher share of students from the townships enrolled based on bursaries and very limited funding. Hence, in the following the neighbourhood wealth of the children, which was merged in from the Census 2011 data, was used to calculate the school average learner wealth for the CEMIS learner cohort. This new variable should be a good proxy for school quality, given the findings by Case and Deaton (1999) and Yamauchi (2011) that show that school location and total schooling fees explain most of the variation in learning outcomes for South Africa. Finally, five quintiles of school average learner wealth can be constructed, which will give a more accurate indication of the school funding and learner background than the old school quintiles or simply taking the wealth of the neighbourhood the school is based at. For NIDS the sample size by school did not allow for calculating average school wealth quintiles. Therefore, the Census wealth quintiles from the neighbourhood of the school location were directly used to calculate the school quintiles at a national level.

The maps shown in Figure 1 and Figure 2 (and Figure A3 in the Appendix) paint the picture of municipalities where poor and rich communities are in close proximity to each other. They also show the historic placement of well-funded former model C (white) schools in the formal areas close to the town centres. Given the findings of previous studies that most parents would like to send their children to former model C schools (e.g. de Kadt et al., 2014; Msila, 2015) we should observe at least some children from the poorest neighbourhoods commuting to richer former model C schools in town.

[11] Fuzzy matching was done using the user written Stata command "reclink". That means that in cases where no perfect match between the key fields in the two datasets existed, the best match, ranked by a matching score, was manually reviewed.
[12] This data was accessed through the DataFirst Secure Research Data Centre at the University of Cape Town.

Table 3 gives the difference in school average learner wealth and the learner neighbourhood wealth quintiles for the municipality of Cape Town using the information from the CEMIS and Census 2011 data. It shows that learners indeed are commuting to schools different than their own neighbourhood wealth. In 2007, 49% of students went to the schools in the same wealth quintile as the area they resided in, 21% went to a poorer and 30% to a richer school quintile. In 2013, the share of students that went to the same quintile was even less at 43%, and about 40% went to a richer school. [13] [14] Hence, as expected the number of students in the CEMIS cohort going to poor funded and lower quality schools declines from grade 6 in 2007 to grade 12 in 2013 because of higher school dropouts within these schools.

Table 3: Difference in school average learner wealth and learner neighbourhood wealth quintiles (Source: CEMIS and Census 2011 data)

		2007		2013	
School average learner wealth Quintile – learner location wealth quintile	-4	51	0.1%	14	0.1%
	-3	894	2.0%	258	1.0%
	-2	1 660	3.7%	726	2.8%
	-1	6 989	15.4%	3 566	13.8%
	0	21 978	48.5%	10 977	42.6%
	1	8 820	19.5%	6 279	24.4%
	2	3 125	6.9%	2 488	9.7%
	3	1 516	3.4%	1 173	4.6%
	4	245	0.5%	310	1.2%

As stated earlier, both the CEMIS and the NIDS data allow us to calculate the actual distance students travel to school (see Table 4 for Cape Town and Table A3 and A4 for the whole country). It shows that white and black learners travel the furthest on average. However, in 2007 (2013) about 8% (18%) of the black children travelled more than 10km to school every day in comparison to 4% (10%) of white students, causing black children to go the furthest on average. Those finding are in line with previous studies (e.g. Hunter, 2015) showing that middle class white students are actually driving relatively far to go to prestigious good schools in the city, while black students have to travel all the way from the townships to the city centre to access quality education. As expected, given the higher number of primary schools, children travel about twice as far in the secondary than in the primary school phase. Interestingly, if we look at the national statistics in Table A3 and A4 (in the Appendix) rural children have to travel further in the primary school phase, while urban children live away further away from their chosen school in secondary school phase.

Given South Africa's racial as well as spatial segregation in terms of school attendance, Table 5 and Table 6 give the wealth of the school and student location by race group for the whole country. While black

[13] It shows that these numbers are close in line with the averages of the whole country (see Table A4 in the Appendix).

[14] It is possible to observe more students going to richer school quintiles than their own neighbourhood wealth, since the neighbourhood wealth quintiles are based on Census 2011 population weighting.

students predominantly live and go to school in the poorest two neighbourhood quintiles, white children live in the top two quintiles and almost always attend the richest schools. Interestingly, it seems that Indian and to some extent coloured students usually go to schools based in richer areas than their own.[15]

Table 43: Distance children live away from their school in km for the municipality of Cape Town (Source: CEMIS)

	2007			2013		
	Mean	Median	Freq.	Mean	Std. Dev.	Freq.
Black	2.943	0.841	9 891	5.166	1.794	4 911
Coloured	2.097	0.713	23 474	3.361	1.469	14 280
Indian/Asian	2.359	1.252	436	3.477	2.249	287
Other	2.454	0.782	945	3.910	2.870	147
White	2.732	1.536	3 418	4.143	2.300	2 154
Total	2.385	0.841	38 164	3.850	1.794	21 779

Table 5: Students neighbourhood quintile and race

	Quintile 1	**Quintile 2**	**Quintile 3**	**Quintile 4**	**Quintile 5**
Black	50.81	24.37	12.44	9.45	2.93
Coloured	6.73	22.03	30.87	31.13	9.23
Indian	32.79	0	1.64	32.79	32.79
White	0.95	3.81	2.86	15.24	77.14

(Source: NIDS wave 1-3)

Table 6: School quintile and race

	Quintile 1	**Quintile 2**	**Quintile 3**	**Quintile 4**	**Quintile 5**
Black	39.08	24.73	17.24	10.46	8.49
Coloured	1.8	6.1	22.62	33.21	36.27
Indian	0	2.08	2.08	12.5	83.33
White	1.33	0	1.33	5.33	92

(Source: NIDS wave 1-3)

These descriptive findings can be formally tested in a simple OLS model which regresses the difference in the revised school quintile to the learner wealth quintile on race, age and a few household characteristics (as seen in Table A7 in the appendix). It shows that Indian families are more likely to send their children to high quality schools. The same is true for parents with more years of education (column 1), parents with matric (column 2) and households with higher per capita income (column 3). Finally, it appears to be easier for families to send their children to richer schools when they reside in urban areas. These results seem to be in line with other studies on school choice behaviour for South Africa (e.g. Msila, 2005; de Kadt et al.; 2014; Hill, 2015; and Hunter, 2015), providing confidence that the wealth quintiles provide sensible information for further analysis.

[15] Looking at provincial mobility, Table A5 in the Appendix shows that, students from the rural province of the Eastern Cape are slightly less mobile in terms of moving between richer and poorer neighbourhoods than children in the Western Cape.

4 Regression analysis

In this part, the relative importance of family background, neighbourhood effects and school quality will be tested. In a first step, it is determined how well the school neighbourhood performs as a proxy for school quality in comparison to the official school quintiles from the DBE. Second, using this proxy for school quality and wealth, different regressions on education and labour market outcomes are run on the CEMIS and NIDS panel data sets.

4.1 Revised versus official DBE school quintiles

Using a very simple model, we compare the appropriateness of the two school wealth measures predicting the average matric 2014 school results. This should give a first indication on how well the revised school wealth measure "*School neighbourhood quintile*" performs in comparison to the official DBE school quintiles.[16] The results in Table 7 column (1) and column (3) show that for the full sample containing all the schools in the country, both measures seem to do similarly well in explaining the variation in average school results with an R^2 of 34% and 33% respectively. On average, the higher the school quintile the better the school performs. It is worth noting that both measures predict that children enrolled in the top quintile school have 11 percentage points' higher matric examination results than those in the poorest quintile.

In column (2) and (4) the same specifications are estimated for the metropolitan samples of the cities of Cape Town, Ethekwini (Durban), Johannesburg and Pretoria. For schools based in these metropolitan regions, it appears that the current neighbourhood wealth quintile of a school (column 4) has a significantly higher model fit (R^2) than the official school quintiles (column 2). In addition, there seem to be larger effects of going to quintile 4 and 5 for the revised measure than the official one. Of course, by not controlling for household wealth or parent's education there is a large omitted variable bias in this simple model, so one should be cautious of interpreting the coefficient estimates as the causal effects of school resources. That is, children living in wealthy neighbourhoods have richer and better educated parents, which should also influence a child's matric performance. However, it is interesting that the coefficient for the school neighbourhood wealth index decreases by about two-thirds after controlling for the share of SGB-teachers[17], which should give a first indication that this measure is a good proxy for the funding and quality of the school.

[16] The official DBE school quintiles are taken from the master list of schools. The quintiles were developed by the DBE in the late 1990s using information from Census 1996 to rank the schools based on the wealth of their community (Van Wyk, 2015). These quintiles are still highly relevant today since they determine the financial support a school is entitled to by the government. Second, in the absence of other measures they have been used as a proxy for school quality in many studies (e.g. Van der Berg, 2008; Spaull and Kotze, 2015). Our revised school quintile measure has the advantage that it is based on much newer Census 2011 data, which provides more recent income information regarding current wealth in a neighbourhood, particularly in the metropolitan regions.

[17] SGB-teachers are employed by the school governing body (SGB) and not by the government. SGB-teachers might be more motivated to teach than public teachers since they have different contracts and can potentially lose their job. In addition, they show that the school has funding capacities to employ these extra teachers.

Table 7: Matric 2014 average school results and school quintiles

	(1) Full sample	(2) Metro	(3) Full sample	(4) Metro	(5) Metro	(6) Metro
VARIABLES	Matric average school percentage	Matric average school percentage	Matric average school percentage	Matric average school percentage	Matric average school percentage	Matric average school percentage
2. DBE school quintile	1.361***	-0.411				
	(0.229)	(1.041)				
3. DBE school quintile	1.486***	-1.389				
	(0.218)	(0.921)				
4. DBE school quintile	4.325***	0.420				
	(0.254)	(0.907)				
5. DBE school quintile	11.10***	7.811***				
	(0.231)	(0.874)				
2. School neighbourhood quintile			1.740***	0.942		
			(0.207)	(0.685)		
3. School neighbourhood quintile			2.066***	1.095*		
			(0.229)	(0.635)		
4. School neighbourhood quintile			4.353***	4.755***		
			(0.223)	(0.630)		
5. School neighbourhood quintile			11.45***	12.19***		
			(0.221)	(0.642)		
School neighbourhood wealth index					1.405***	0.427***
					(0.0559)	(0.0643)
SGB teacher share						31.58***
						(1.456)
Constant	45.56***	48.32***	45.56***	46.91***	47.79***	47.44***
	(0.167)	(0.823)	(0.140)	(0.524)	(0.227)	(0.187)
Observations	5 996	989	5 996	989	989	986
R-squared	0.342	0.323	0.334	0.401	0.390	0.588

Standard errors in parentheses, *** p<0.01, ** p<0.05, * p<0.1;
Number of matric students used as weights (Source: DBE school data)

4.2 CEMIS regression results

In Table 8 the regression results for the likelihood of each grade 6 student reaching and passing matric without repeating class are presented. First, it is tested which wealth variable predicts the student success best. For the "old" school quintiles[18] in regression column (1) the first two quintiles were added together since hardly any quintile 1 school exist in Cape Town. It shows that if a child goes to a quintile 5 school he/she will be 13% more likely to reach and pass matric in 2013 than a child in a quintile 1 or 2 school. Surprisingly, students from quintile 3 and 4 schools are actually less likely to reach matric than from quintile 1 and 2 schools, which indicates the need to update the old school quintiles currently used by the DBE. On the other hand, the neighbourhood quintiles of the learner's household location (column 2) and of the school's location (column 3) as well as the school average learner wealth quintile (column 4) have significant coefficients with the expected sign. That is, each increase in wealth quintiles of the

[18] The "old" school quintiles are the official quintiles from the master list of schools provided by DBE.

neighbourhood or of the school, improves the student's likelihood of reaching matric without repetition after grade 6 and then also passing matric. However, it also shows that the large effects are only seen for the richest school average quintile.

Table 8: OLS regression – reaching and passing matric without repeating (Source: CEMIS)

VARIABLES	(1) Matriculated	(2) Matriculated	(3) Matriculated	(4) Matriculated	(5) Matriculated
Coloured	0.0228***	0.0246***	0.0288***	0.00253	0.000728
	(0.00684)	(0.00506)	(0.00516)	(0.00527)	(0.00529)
Indian	0.150***	0.126***	0.129***	0.0554***	0.0426**
	(0.0218)	(0.0215)	(0.0213)	(0.0215)	(0.0215)
White	0.150***	0.0970***	0.0862***	0.00887	-0.00895
	(0.0100)	(0.0110)	(0.0103)	(0.0106)	(0.0116)
Overage (in gr 6)	-0.267***	-0.272***	-0.265***	-0.262***	-0.262***
	(0.00495)	(0.00493)	(0.00493)	(0.00491)	(0.00490)
Underage (in gr 6)	0.0460***	0.0431***	0.0496***	0.0538***	0.0533***
	(0.00937)	(0.00936)	(0.00933)	(0.00929)	(0.00928)
2.Learner neighbourhood quintile		-0.00391			-0.00157
		(0.00699)			(0.00694)
3.Learner neighbourhood quintile		0.0333***			0.0143**
		(0.00652)			(0.00706)
4.Learner neighbourhood quintile		0.162***			0.0590***
		(0.00696)			(0.00821)
5.Learner neighbourhood quintile		0.227***			0.0764***
		(0.0102)			(0.0116)
3.Old school quintile	-0.0407***				
	(0.00807)				
4.Old school quintile	-0.0240***				
	(0.00883)				
5.Old school quintile	0.138***				
	(0.00872)				
2.School neighbourhood quintile			0.0111		
			(0.00757)		
3.School neighbourhood quintile			0.0174**		
			(0.00682)		
4.School neighbourhood quintile			0.158***		
			(0.00696)		
5.School neighbourhood quintile			0.253***		
			(0.00900)		
2.School average learner quintile				0.00598	-0.00306
				(0.00637)	(0.00685)
3.School average learner quintile				0.0602***	0.0412***
				(0.00716)	(0.00786)
4.School average learner quintile				0.149***	0.114***
				(0.00710)	(0.00839)
5.School average learner quintile				0.310***	0.257***
				(0.00823)	(0.0102)
Constant	0.370***	0.346***	0.329***	0.337***	0.335***
	(0.00619)	(0.00587)	(0.00600)	(0.00531)	(0.00616)
Observations	44 277	44 277	44 277	44 277	44 277
R-squared	0.138	0.137	0.143	0.151	0.153

Standard errors in parentheses, *** p<0.01, ** p<0.05, * p<0.1

That is, if a student in grade 6 is going to one of the richest 20% of schools in terms of average learner wealth, he is 31% more likely to pass matric than students going to one of the 20% poorest school. The large school effect holds true even after controlling for student neighbourhood wealth at the same time in column (5). Looking at the size of the coefficients it seems that the school a child is attending and not his

or her home neighbourhood is driving most of the differences in schooling outcomes. In addition, the average school wealth variable has the highest predictive power of all the wealth measures. Hence, in further specifications the school learner average quintile, instead of simple school location wealth quintile, will be used.

It is worth noting that being overaged in grade 6 reduces the likelihood of passing matric highly significant and by about 27%. That is, failing a grade once in the first 6 years (and thus being overaged in grade 6) seems to massively decrease the chances of ever reaching and passing matric, which is in line with previous studies on Cape Town and national data (e.g. Lam et al. 2011, Branson et al. 2015). In terms of racial effects, they only seem to play an important role in the first 3 columns. As soon as one controls for the average learner wealth at a school, the coloured and white dummy variables turn insignificant. This implies that black students are not less likely to pass matric on time, that is, if they manage to go to a well-functioning school. However, as noted before, only 5% of all black students can actually afford to go to the richest 20% of schools in Cape Town.

Table A8 in the Appendix shows the same regression model for the sample of students that have point GPS coordinates to calculate their exact distance to school. In column (1) and (2) the log(distance) to school variable is significant and positive, indicating that students driving further to school are more likely to reach and pass matric in time. This variable could be a proxy for parent's motivation and involvement to send their child to a quality school. Yet, after controlling directly for school quality in column (3-4), the distance variable turns insignificant.

The next outcome variable is the standardized grade 9 test result from Western Cape Systemic. Table 9 columns (1) and (2) show the regression results for language and math. It appears that there is a larger effect being in one of the richest 20% of schools and living in the richest part of town for mathematics than for language test results. This and the larger coefficient for being a white student in column (2) seems to be in line with national test results like TIMSS (see Ready 2011). Overall, the difference between children going to the poorest vs. the richest 20% schools is about 1.3 standard deviation, while living in the richest parts of town results would add another 0.4 standard deviation higher test results. In total, that would imply a learning gap between the poorest and richest children, going to the best and worst schools of metropolitan Cape Town, of about 1.7 standard deviations, which translates to approximately 5 years' worth of learning, using Spaull and Kotze's (2015) recommendation that 0.3 standard deviations could be considered approximately equivalent to one year worth of learning. Moreover, the larger difference in mathematics test results is in particularly alarming in terms of the labour market perspectives of students from poor background, given that higher grade mathematics is a requirement for university entry for many courses in sciences, engineering and commerce (Van der Berg 2009).

Table 9: OLS regression – grade 9 exam results (Source: CEMIS and Western Cape Systemic Tests)

VARIABLES	(1) Std (gr9) language	(2) Std (gr9) math	(3) Std (gr9) aggregate	(4) Std (gr9) aggregate
Coloured	0.165***	0.0608***	0.113***	-0.00170
	(0.0148)	(0.0131)	(0.0130)	(0.0129)
Indian	0.466***	0.502***	0.532***	0.567***
	(0.0474)	(0.0422)	(0.0417)	(0.0429)
White	0.250***	0.539***	0.465***	0.571***
	(0.0272)	(0.0242)	(0.0239)	(0.0242)
Age in years	-0.335***	-0.188***	-0.267***	-0.269***
	(0.00806)	(0.00718)	(0.00709)	(0.00731)
Ln(Distance to school)	0.0689***	0.0434***	0.0584***	0.0562***
	(0.00442)	(0.00393)	(0.00388)	(0.00400)
2.Learner neighbourhood quintile	0.0565***	0.0225	0.0383**	0.328***
	(0.0189)	(0.0168)	(0.0166)	(0.0178)
3.Learner neighbourhood quintile	0.0797***	0.0267	0.0519***	0.569***
	(0.0183)	(0.0163)	(0.0161)	(0.0177)
4.Learner neighbourhood quintile	0.183***	0.138***	0.170***	1.051***
	(0.0193)	(0.0172)	(0.0170)	(0.0185)
5.Learner neighbourhood quintile	0.302***	0.435***	0.417***	1.748***
	(0.0264)	(0.0235)	(0.0232)	(0.0249)
2.School average learner quintile	0.119***	0.00586	0.0539***	
	(0.0187)	(0.0167)	(0.0165)	
3.School average learner quintile	0.311***	0.0864***	0.189***	
	(0.0201)	(0.0178)	(0.0176)	
4.School average learner quintile	0.670***	0.364***	0.527***	
	(0.0215)	(0.0191)	(0.0189)	
5.School average learner quintile	1.092***	1.273***	1.304***	
	(0.0255)	(0.0227)	(0.0224)	
School learner average wealth - learner wealth quintile				0.303***
				(0.00527)
Constant	4.342***	2.266***	3.348***	3.251***
	(0.127)	(0.113)	(0.111)	(0.115)
Observations	24 219	24 266	24 182	24 182
R-squared	0.393	0.542	0.546	0.516

Standard errors in parentheses, *** p<0.01, ** p<0.05, * p<0.1

The finding that South Africa has basically two education systems, one for the poor (black and coloured) students attending formerly disadvantaged schools and one for the well-off, that produces vastly different schooling results, have been shown by many other studies in the past e.g. Van der Berg (2007, 2008), Reddy (2011) and Spaull and Kotze (2015). Yet, the question remains: what is the effect of coming from a poor neighbourhood while attending a richer and higher quality school? The results provided in column (4) of Table 9, show that for each quintile of school wealth higher than the student's neighbourhood wealth, the test results improve by 0.3 standard deviations, after controlling for student neighbourhood wealth and school to distance. This implies that a student living in the poorest quintile neighbourhood, going to a quintile 5 wealth school, would perform 1.2 standard deviation better than a student from the same neighbourhood attending a school one of the poorest 20% of schools, or approximately 4 years' worth of learning. That implies that about 70% of the learning gap between the richest and the poorest students is overcome if they attend the same quality school. However, as mentioned before, due to financial and transport constraints only 3% of children living in the poorest 60% of Cape Town municipality neighbourhoods are actually manging to go to the highest school quintile.

In terms of the other variables, as before, older students perform significantly worse and the distance to school, which can be seen as a proxy for parent's involvement, has a significant and positive coefficient in all specifications. In a further step, the same specification can be run on the standardized average matric results in grade 12. Overall, the results seem to be in line with the earlier findings, even though the effect of going to one of the richest 20% is slightly smaller now (see Table A9 in the Appendix). However, it is more difficult to interpret the results, since those students that can be observed in grade 12 are significantly different from the average youth in Cape Town who is no longer enrolled in school, due to the high number of school drop outs between grade 9 and grade 12 (also see e.g. Spaull and Kotze 2015, Van Wyk 2015). Second, information on average matric test results was only available for about half of the grade 12 sample.

One problem when interpreting the large impact of the school average wealth variable on schooling outcomes for the poor, is unobserved heterogeneity. That is, one could argue children travelling into town have vastly different parents than those staying in the same neighbourhood but going to a township school. Hence, there might be some unobserved heterogeneity which is not be captured by the distance to school variable. Therefore, in the next section the robustness of the results will be tested by using the NIDS panel, which allows controlling for household income and parents education while still observing students movement between neighbourhoods.

4.3 NIDS short-run education effects

The advantage of using NIDS is that it not only provides the possibility to test long-term outcomes of quality education but also that it is representative at a national level. Table 10 reports the pooled OLS regression results from NIDS wave 1-3, with years of education reached as the dependent variable for the sub-group of 15-18 year olds.[19] This age group is of particular interest since most grade repetition and school drop-outs occur in grades 9-12 (van Wyk, 2015). All regressions use population weights to adjust for attrition and clustered standard errors. In case the youth were observed when they had already left the school system, the information of the last school before the dropout has been linked to the individual.

Column (1) of Table 10 shows the simple model with just age and the revised school neighbourhood quintiles as explanatory variables. In this specification children going to the richest school quintile, quintile 5, have attained roughly one more year of education than the children from the poorest schools at the same age. As we start controlling for race and gender in column (2), parental education in column (3) and household income in column (4) the coefficient drops to about one-quarter but remains sizable and statistically significant. There is a high correlation between parental education, household income and school wealth quintiles.

[19] Table A10 in the Appendix shows the average household characteristics for age 15-18. Note the high percentage of fathers missing, the low mean per capita income and average years of education for parents.

Table 10: Pooled OLS regression: reached years of education by age 15-18

VARIABLES	(1) Years of education	(2) Years of education	(3) Years of education	(4) Years of education
Age	0.720***	0.708***	0.729***	0.725***
	(0.0225)	(0.0210)	(0.0208)	(0.0205)
Male		-0.520***	-0.517***	-0.519***
		(0.0790)	(0.0779)	(0.0788)
White		0.299*	-0.178	-0.327*
		(0.176)	(0.166)	(0.171)
Indian		0.848***	0.425**	0.297
		(0.214)	(0.187)	(0.190)
Coloured		0.266**	0.152	0.0839
		(0.117)	(0.108)	(0.113)
Mother's years of education			0.0659***	0.0604***
			(0.00860)	(0.00865)
Mother not in the household			-0.161**	-0.173**
			(0.0791)	(0.0786)
Father's years of education			0.0455***	0.0380***
			(0.00787)	(0.00786)
Father not in the household			-0.133**	-0.110**
			(0.0514)	(0.0500)
Ln(per capita income)				0.119***
				(0.0262)
2. School neighbourhood quintile	0.255**	0.175	0.0786	0.0745
	(0.126)	(0.111)	(0.106)	(0.106)
3. School neighbourhood quintile	0.425***	0.323***	0.149	0.139
	(0.131)	(0.101)	(0.0985)	(0.0985)
4. School neighbourhood quintile	0.618***	0.414***	0.166	0.137
	(0.125)	(0.120)	(0.113)	(0.113)
5. School neighbourhood quintile	1.029***	0.715***	0.312***	0.258**
	(0.108)	(0.115)	(0.119)	(0.121)
Constant	-3.598***	-3.173***	-3.980***	-4.524***
	(0.360)	(0.368)	(0.360)	(0.385)
Observations	7 254	7 254	7 247	7 245
R-squared	0.267	0.314	0.357	0.360

Not reported province dummies. Standard errors in parentheses; *** p<0.01, ** p<0.05, * p<0.1.
(Source NIDS wave 1-3)

If we estimate the same specifications but replacing the revised school quintiles with the DBE school quintiles (see Table A11 in the Appendix), the coefficients for school quintile dummies become insignificant and negligible in magnitude once we control for parental education and family income. Hence, it seems that the revised school wealth quintiles provide a more informative measure of school quality than the official DBE school quintile measure. One obvious issue with using years of education obtained is that this does not necessarily imply actual learning but merely years of schooling completed. The true learning gap between the poorest and richest schools may be even larger. [20]

[20] Estimating the same model on an administrative data set for learners from Cape Town and using Western Cape Systemic Tests of learner performance shows that the learning gap between the poorest and richest children, going to the best and worst schools of Cape Town, is about 1.7 standard deviations, which translates to approximately 5 years' worth of learning. This demonstrates

It is also noteworthy that there appears to be a large and significant gender effect, causing boys to obtain about half a year less education at the same age as girls. Interestingly, there seem to be no remaining substantial race effect after controlling for household income, parental education and school characteristics. This means black children have the same grade progression if they have similar socio-economic backgrounds and attend the same schools as children from the other races. Yet, the average black child has a mother and father with about 7 years of education and lives in a household with a mean per capita income of about R900, whereas the average white child has parents with about 13 years of education and R7000 mean household income. Taking all of this into consideration, an average black child has reached about 1.5 years less education at the same age than the average white child. In addition, the outcome variable only measures school attainment despite differences in the quality of education the average black and white child receives.

Column (1) of Table 11 controls for the neighbourhood wealth quintile and the distance children travel to school. While the neighbourhood the child lives in does not seem to have an additional significant effect, the distance to school is a positive and significant determinant of schooling attainment. This could be either because parents' who send their children to schools further away are also more likely to support their offspring with their education in other ways, or because parents are aware of differences in school quality that are unobservable to the econometrician.

To control for other sources of unobserved heterogeneity between neighbourhoods, column (2)-(5) adds cluster fixed effects to the model. In column (2) the coefficient for the highest school quintile does increase to 0.35 from 0.28. The same specification is estimated for the urban sample of NIDS youth in column (3). As seen before in Table 10, the school wealth quintile variable seems to explain education outcomes particularly well in South African cities. However, the significant positive coefficient for the "difference in school to household quintile" in column (4) implies that students living in the poorest quintile neighbourhood but going to a quintile 5 wealth school, would reach about 0.6 year more education than a students from the same neighbourhood who attend one of the poorest 20% of schools.

For the rural sample in column (5) no positive effect for such movement between neighbourhoods can be observed. Yet here there appears to be large gains to attending schools further away from home. Finally, in Table A8 in the appendix, some additional school quality variables like the share of SGB-teachers at a school, dummies for private-or mixed funded schools or the student-teacher ratio are added to the model. The coefficient of the school wealth quintile variable stays robustly positive and significant.

---

that the school neighbourhood effect is perhaps even more important in determining learning than its effect on schooling attainment.

Table 11: Pooled OLS and cluster FE regressions: reached years of education by age 15-18

	(1) Full sample	(2) Full sample	(3) Urban sample	(4) Urban sample	(5) Rural sample
VARIABLES	**OLS**	**Cluster FE**	**Cluster FE**	**Cluster FE**	**Cluster FE**
Age	0.705***	0.699***	0.747***	0.748***	0.651***
	(0.0223)	(0.0227)	(0.0349)	(0.0352)	(0.0293)
Male	-0.493***	-0.487***	-0.270**	-0.268**	-0.727***
	(0.0767)	(0.0806)	(0.108)	(0.109)	(0.0790)
White	-0.403**	0.0253	-0.0220	-0.0246	-0.654
	(0.177)	(0.249)	(0.262)	(0.261)	(0.809)
Indian	0.218	0.252	0.124	0.144	0.895***
	(0.195)	(0.253)	(0.300)	(0.311)	(0.168)
Coloured	0.154	0.356	0.389	0.383	-1.158
	(0.111)	(0.267)	(0.260)	(0.260)	(0.809)
Ln (School distance)	0.0374*	0.0684**	0.0217	0.0144	0.148***
	(0.0226)	(0.0272)	(0.0388)	(0.0380)	(0.0345)
Ln(per capita income)	0.107***	0.0784**	0.0811*	0.0810*	0.0895**
	(0.0318)	(0.0306)	(0.0479)	(0.0480)	(0.0422)
Mother's years of education	0.0628***	0.0586***	0.0533***	0.0545***	0.0649***
	(0.00934)	(0.00987)	(0.0189)	(0.0197)	(0.0105)
Mother not in the household	-0.150**	-0.132*	-0.371***	-0.375***	0.0563
	(0.0743)	(0.0732)	(0.130)	(0.130)	(0.0791)
Father's years of education	0.0366***	0.0293***	0.0307**	0.0298**	0.0303***
	(0.00853)	(0.00801)	(0.0146)	(0.0149)	(0.00983)
Father not in the household	-0.110**	-0.0965**	-0.0236	-0.0254	-0.161***
	(0.0486)	(0.0487)	(0.0812)	(0.0799)	(0.0540)
Household neighbourhood wealth quintile	YES	YES	YES	YES	YES
2. School neighbourhood quintile	0.0810	-0.117	0.222		
	(0.113)	(0.134)	(0.222)		
3. School neighbourhood quintile	0.158	0.219	0.527**		
	(0.112)	(0.168)	(0.237)		
4. School neighbourhood quintile	0.127	0.189	0.544**		
	(0.125)	(0.171)	(0.222)		
5. School neighbourhood quintile	0.277*	0.354**	0.689***		
	(0.141)	(0.169)	(0.205)		
Difference school to household quintiles				0.155***	-0.0681
				(0.0519)	(0.0753)
Constant	-4.213***	-4.400***	-5.512***	-5.425***	-2.913***
	(0.396)	(0.468)	(0.606)	(0.629)	(0.974)
Observations	7 131	7 131	2 716	2 714	4 415
R-squared	0.354	0.453	0.515	0.514	0.405

Standard erros in parentheses; *** p<0.01, ** p<0.05, * p<0.1.
Not reported province dummies. (Source NIDS wave 1-3)

4.4 NIDS long-run education effects

We now consider the effect of quality schooling on university enrolment and labour market earnings. In Table 12 university enrolment is regressed on race, age and household characteristics as well as school and household neighbourhood quintiles for the age group 18-23. The counterfactual group for those enrolled at university are youth directly starting to work after school, unemployed or economically not-active. The NIDS panel element was used, to get the actual information from the last school attended and the neighbourhood the youth was living in whilst still going to school. In this model, household wealth and

coming from the richest neighbourhood quintile seem to substantially increase the chances of a youth enrolling in university. This finding suggests some kind of credit constraints for poorer students going to university, which is not surprising given the high direct costs of university fees. There also seem to be a significantly positive effect from the school quality measure, particularly for those from the highest school quintile. Black youth seem to be more likely to start some form of tertiary education after controlling for social-economic status.

Table 12: Pooled OLS regression: youth enrolled in university age 18-23

	(1) Full sample	(2) Full sample	(3) Full sample	(4) Urban sample
VARIABLES	**University enrolment**	**University enrolment**	**University enrolment**	**University enrolment**
Age	0.0234***	0.0267***	0.0292***	0.0438***
	(0.00584)	(0.00551)	(0.00583)	(0.00875)
Male	-0.0463***	-0.0553***	-0.0652***	-0.0697***
	(0.0135)	(0.0128)	(0.0136)	(0.0202)
White	0.293***	-0.0333	-0.189***	-0.205***
	(0.0301)	(0.0326)	(0.0388)	(0.0464)
Indian	-0.0223	-0.283***	-0.400***	-0.480***
	(0.0495)	(0.0485)	(0.0506)	(0.0707)
Coloured	-0.0385	-0.140***	-0.197***	-0.210***
	(0.0285)	(0.0274)	(0.0299)	(0.0357)
Ln(per capita income)		0.0928***	0.0831***	0.114***
		(0.00729)	(0.00802)	(0.0124)
Mother's years of education		0.0147***	0.0134***	0.0165***
		(0.00189)	(0.00205)	(0.00326)
Father's years of education		0.00962***	0.00701***	0.00226
		(0.00203)	(0.00220)	(0.00344)
2.Ex School neighbourhood quintile			-0.0114	0.0740*
			(0.0214)	(0.0431)
3. Ex School neighbourhood quintile			-0.0101	-0.0179
			(0.0257)	(0.0444)
4. Ex School neighbourhood quintile			0.0403	0.0437
			(0.0263)	(0.0436)
5. Ex School neighbourhood quintile			0.123***	0.115***
			(0.0270)	(0.0428)
2.Ex household neighbourhood quintile			-0.0392*	-0.0264
			(0.0208)	(0.0388)
3. Ex household neighbourhood quintile			-0.0367	0.0103
			(0.0263)	(0.0428)
4. Ex household neighbourhood quintile			-0.0391	-2.43e-05
			(0.0267)	(0.0424)
5. Ex household neighbourhood quintile			0.124***	0.121**
			(0.0351)	(0.0503)
Constant	-0.358***	-1.164***	-1.076***	-1.575***
	(0.122)	(0.123)	(0.130)	(0.204)
Observations	3 443	3 440	3 095	1 539
R-squared	0.068	0.176	0.202	0.263

Standard errors in parentheses; *** p<0.01, ** p<0.05, * p<0.1; (Source: NIDS wave 1-4)

To identify the effect of school quality on earnings, we estimate several earnings functions using OLS in Table 13. The dependent variable is log earnings and the control variables include gender, race and age[21]. The variables of interest are years of education, years of education squared and the wealth index of the school the young worker went to as a measure of school quality. The information for the school was taken from the retrospective question of NIDS that asked which school a respondent last attended. Only the subgroup of workers aged 20-30 were chosen, since there are concerns around the quality of the merge on school names for older cohorts. In addition, since the school quintiles seem to explain the variation in education outcomes best in urban settings, the analysis was further limited to this subgroup.

Table 13: Pooled OLS regression: ln(earnings) of age 21-30 in urban sample

	(1) Full sample	(2) Full sample	(3) Black sample	(4) Black sample
VARIABLES	**Ln(earnings)**	**Ln(earnings)**	**Ln(earnings)**	**Ln(earnings)**
White	0.286***	0.277***		
	(0.104)	(0.104)		
Coloured	0.0648	0.0711		
	(0.0541)	(0.0541)		
Indian	0.652***	0.623***		
	(0.162)	(0.163)		
Age	0.0575***	0.0579***	0.0564***	0.0568***
	(0.00647)	(0.00647)	(0.00757)	(0.00755)
Male	0.264***	0.265***	0.257***	0.259***
	(0.0323)	(0.0323)	(0.0376)	(0.0375)
Mother having matric	0.185***	0.175***	0.234***	0.216***
	(0.0530)	(0.0531)	(0.0616)	(0.0616)
Father having matric	0.173***	0.166***	0.168**	0.165**
	(0.0554)	(0.0554)	(0.0673)	(0.0671)
Education	-0.143***	-0.131***	-0.166***	-0.151***
	(0.0390)	(0.0393)	(0.0439)	(0.0440)
Education^2	0.0157***	0.0148***	0.0169***	0.0161***
	(0.00186)	(0.00190)	(0.00213)	(0.00214)
Std (school index)	0.0429**	-0.187**	0.0616***	-0.394***
	(0.0181)	(0.0951)	(0.0203)	(0.122)
Std (school index)* education		0.0204**		0.0403***
		(0.00827)		(0.0106)
Constant	5.318***	5.277***	5.489***	5.373***
	(0.277)	(0.277)	(0.314)	(0.315)
Observations	3 003	3 003	2 254	2 254
R-squared	0.254	0.256	0.236	0.241

Standard errors in parentheses; *** p<0.01, ** p<0.05, * p<0.1. (Source: NIDS wave 1-4)

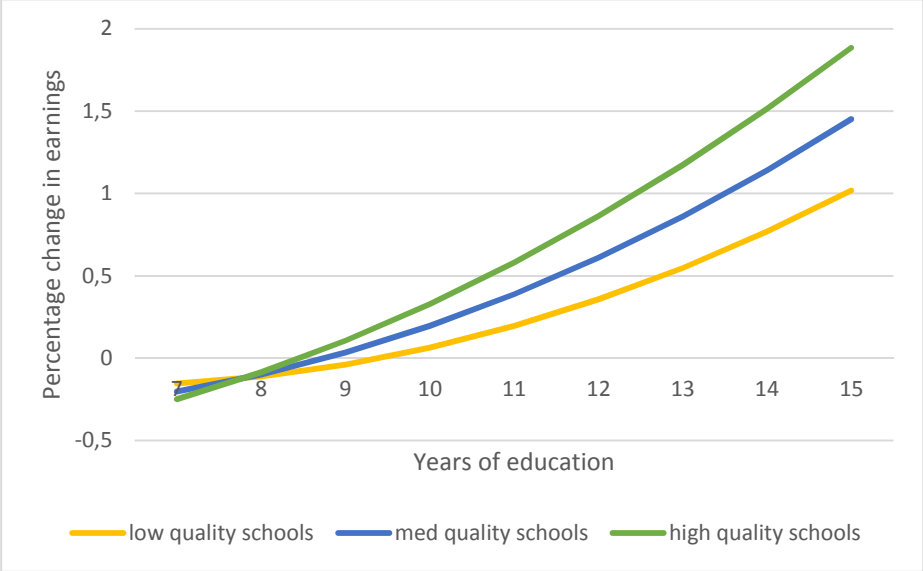
To quantify the impact of education quality on earnings, it is important to distinguish between the direct and indirect effects: the indirect effect allows a student from a better school to attain more years of education (as seen in Table 10), whereas the direct effect is the benefit after controlling for years of

[21] Even though mincer wage regressions normally assume non-linear returns to age, for the small age-period at hand, the assumption of linearity for age seems to be sufficient.

education. The channel this would have to work through is higher ability, or better writing, mathematics or other skills that can be observed by their employer. Lastly there is the potential problem of unobserved household heterogeneity, since children that went to richer schools might also have higher ability, richer parents and other unobserved factors that are financially remunerated in the labour market. In an attempt to control for this household effects, mother and father education are included in the earnings model.

Column (1) in Table 13 shows that school wealth does have a positive and significant effect on earnings, indicating some direct positive effect of quality schooling on earnings. Returns to education seem to be convex given the significant and negative coefficient for education and positive and significant coefficient for education squared. There seem to be a wage premium for being male, Indian[22] and white. Next, to observe if there is a premium for each year higher quality education received, an interaction term between the school wealth index and years of education reached is entered in column (2). Entering this interaction term, the coefficient of education squared marginally decreases and the coefficient for the school wealth index turns negative. Given the significant and positive interaction term this signals that there is a wage premium for higher quality education only when a certain combination of quality and years of education is reached. This is best illustrated in a graph as shown in Figure 3. The graph shows the returns of education for low, middle and high quality schools. For all schools the returns to education turn positive around 8 years.[23] The more years a student reached, the higher the premium for quality education they received. That means that having matric from a high quality school would increase earnings by about 50 percentage points on top of the normal returns to education.

Figure 3: Returns to education given different school quality

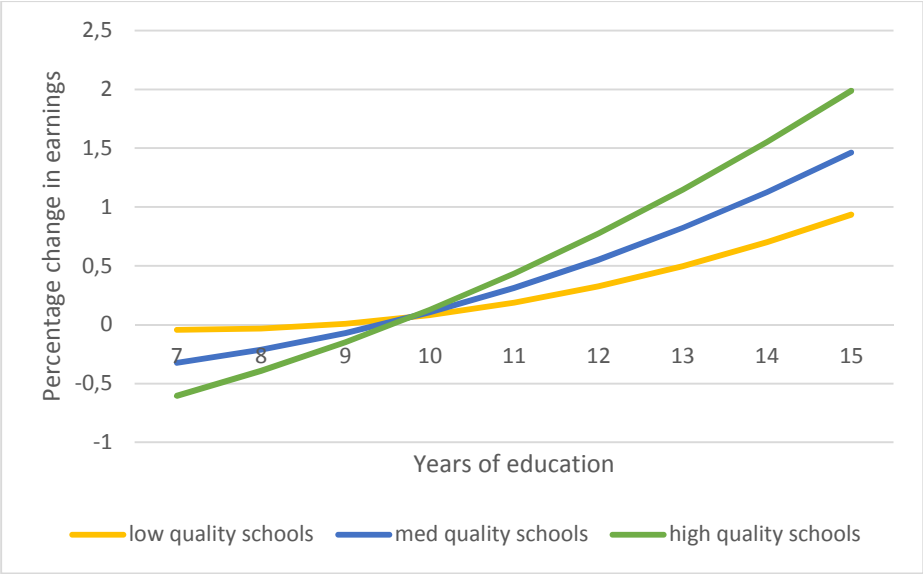


(Source: own calculation from earnings regression seen in Table 8)

[22] The particularly large coefficient for the Indian dummy might be explained by the small number of Indian youth in the sample.
[23] Given that only 4% of workers in our sample have fewer than 8 years of schooling, the model fit at these lower schooling values is not of great practical importance.

As a robustness check, the same set of specifications was also run only on the black sample. It appears that for black youth in urban areas the direct effect of going to a high quality school is even larger than the overall average. This is demonstrated in Figure 4 that has even larger returns to high levels of education for black workers that go to the highest quality schools.

Figure 4: Returns to education given different school quality for black youth



(Source: own calculation from earnings regression seen in Table 8)

5 Conclusion

This paper evaluated the relative importance of family, neighbourhood and school quality in explaining the variation in school and labour market outcomes for South Africa. To do so, the Census 2011 community data was used to develop a new wealth index for about 85 000 neighbourhoods that got linked to school and student location of administrative records CEMIS and the national panel data NIDS. Revised school wealth quintiles were estimated. They prove to be good measures for school quality and are more accurate than the previously used school quintiles in explaining schooling outcomes, in particular for urban areas.

The results of this study have shown that children attending the richest 20% of schools in grade 6 will be 30% more likely to reach and pass matric in time, in comparison to children from the poorest 20% of schools. By grade 9 the learning gap between the poorest and richest children, going to the best and worst schools of Cape Town, would total to about 1.7 standard deviations or approximately 5 years' worth of learning. These findings are in line with the national data showing that a student living in the poorest quintile neighbourhood who attends the richest school in town would perform significantly better than a student from the same neighbourhood who is going to the poorest school. Sadly, due to financial and transport constraints, in South Africa only 10% of children living in the poorest 60% neighbourhoods are managing to attend schools in the richest school quintile. Hence, the study provides further evidence of

how segregated the South African school system is: providing quality education to a few and leaving behind a large share of poor children. It highlights the importance of bringing high quality schooling into the townships and rural areas where a majority of the poor live.

This paper is amongst the first studies to provide evidence of the importance of quality education in explaining who is enrolling in university (after reaching matric) and determining subsequent earnings. Both of these outcomes are of great concern for decreasing inequality in South Africa. From the earnings model it can be concluded that there are increasing returns to education for attending a high quality school. This means that children receive a 50% premium for receiving matric from a high quality school in comparison to a child with matric from a low quality school.

This study has shown that the revised school wealth quintiles are an important means of identifying schools that need special attention, since it predicts individual learner and school outcomes more accurately than the official school quintiles. Such an instrument could also be used by the Department of Basic Education to target poor schools and reform a funding system to achieve more equal school outcomes in South Africa in the long run.

6 Literature

Altonji, J. G., & Mansfield, R. K. (2011). The role of family, school, and community characteristics in inequality in education and labor market outcomes. *Whither opportunity*, 339-358.

Björklund A, Jäntti M, Raaum O, Österbacka E, Eriksson T (2004) Family structure and labor market success: the influence of siblings and birth order on the earnings of young adults in Norway, Finland, and Sweden. In: Corak M (ed) Generational mobility in North America and Europe. *Cambridge University Press*, Cambridge, pp 208–225.

Branson, N., Hofmeyr, C., & Lam, D. (2014). Progress through school and the determinants of school dropout in South Africa. *Development Southern Africa*, 31(1), 106-126.

Boccanfuso, D., Larouche, A., & Trandafir, M. (2015). Quality of higher education and the labor market in developing countries: Evidence from an education reform in Senegal. *World Development*, *74*, 412-424.

Card, D., & Krueger, A. (1990). Does school quality matter? Returns to education and the characteristics of public schools in the United States (No. w3358). National Bureau of Economic Research.

Case, A., & Deaton, A. (1999). School inputs and educational outcomes in South Africa. *Quarterly journal of economics*, 1047-1084.

Case, A., & Yogo, M. (1999). Does school quality matter? Returns to education and the characteristics of schools in South Africa (No. w7399). National Bureau of Economic Research.

de Kadt, J., Norris, S. A., Fleisch, B., Richter, L., & Alvanides, S. (2014). Children's daily travel to school in Johannesburg-Soweto, South Africa: geography and school choice in the Birth to Twenty cohort study. *Children's Geographies*, 12(2), 170-188.

Eriksson T, Zhang Y (2012) The role of family background for earnings in rural China. *Front Econ China*, 7(3):465–477.

Glewwe, P., & Kremer, M. (2006). Schools, teachers, and education outcomes in developing countries. *Handbook of the Economics of Education*, 2, 945-1017.

Hanushek, E. A., Lavy, V., & Hitomi, K. (2008). Do students care about school quality? Developing countries. *Journal of Human Capital*, 2(1), 69–105.

Harbison, R. W., & Hanushek, E. A. (1992). Educational performance of the poor: Lessons from rural northeast Brazil. Oxford University Press.

Hill, L. D. (2016). Race, school choice and transfers to opportunity: implications for educational stratification in South Africa. *British Journal of Sociology of Education*, 37(4), 520-547.

Hunter, M. (2015). Schooling choice in South Africa: The limits of qualifications and the politics of race, class and symbolic power. *International Journal of Educational Development*, 43, 41-50.

Lam, D., Ardington, C., & Leibbrandt, M. (2011). Schooling as a lottery: Racial differences in school advancement in urban South Africa. *Journal of Development Economics*, 95(2), 121-136.
Lindahl, L. (2011). A comparison of family and neighborhood effects on grades, test scores, educational attainment and income—evidence from Sweden. *The Journal of Economic Inequality*, 9(2), 207-226.

Msila, V. 2005. The education exodus: The flight from township schools. *Africa Education Review*, 2 (2): 173–188.

Nicoletti, C., & Rabe, B. (2013). Inequality in Pupils' Test Scores: How Much do Family, Sibling Type and Neighbourhood Matter?. *Economica*, 80(318), 197-218.

Raaum, O., Salvanes, K. G., & Sørensen, E. Ø. (2006). The neighbourhood is not what it used to be*. *The Economic Journal*, *116*(508), 200-222.

Reddy, V. (2012). Towards Equity and Excellence: Highlights from TIMSS 2011. The south african perspective.

Schnitzlein, D. D. (2014). How important is the family? Evidence from sibling correlations in permanent earnings in the USA, Germany, and Denmark. *Journal of Population Economics*, *27*(1), 69-89.

Solon, G., Page, M. E., & Duncan, G. J. (2000). Correlations between neighboring children in their subsequent educational attainment. *Review of Economics and Statistics*, 82(3), 383-392.

Spaull, N., & Kotze, J. (2015). Starting behind and staying behind in South Africa: The case of insurmountable learning deficits in mathematics. *International Journal of Educational Development*, 41, 13-24.

Statistics South Africa, 2012a: Census 2011 Metadata. Statistics South Africa. Pretoria.

Statistics South Africa, 2012b: Census 2011 Statistical release – P0301.4. Statistics South Africa. Pretoria.

Statistics South Africa, 2014: The South African MPI: Creating a multidimensional poverty index using census data. Statistics South Africa. Pretoria.

Taylor, N., Van der Berg, S., & Mabogoane, T. (Eds.). (2013). *Creating effective schools*. Cape Town: Pearson.

Taylor, S., van der Berg, S., Reddy, V., & Janse van Rensburg, D. (2011). How well do South African schools convert grade 8 achievement into matric outcomes? (No. 13/2011). Stellenbosch University, Department of Economics.

Van der Berg, S., Burger, C., Burger, R., de Vos, M., du Rand, G., Gustafsson, M., ... & van Broekhuizen, H. (2011). Low quality education as a poverty trap. Stellenbosch: Stellenbosch University.

Van der Berg, S. (2007). Apartheid's enduring legacy: Inequalities in education. *Journal of African Economies*, 16(5), 849-880.

Van der Berg, S. (2008). How effective are poor schools? Poverty and educational outcomes in South Africa. *Studies in Educational Evaluation*, 34(3), 145-154.

Van der Berg, S. (2009). The persistence of inequalities in education. In B. Kahn & G. Kingdon (Eds.), South African economic policy under democracy (pp. 327–354). Oxford: Oxford University Press

Van Wyk, C. (2015). An overview of education data in South Africa: An inventory approach'. *South African Journal of Childhood Education (Forthcoming)*.

Yamauchi, F. (2011). School quality, clustering and government subsidy in post-apartheid South Africa. *Economics of Education Review*, 30(1), 146-156.

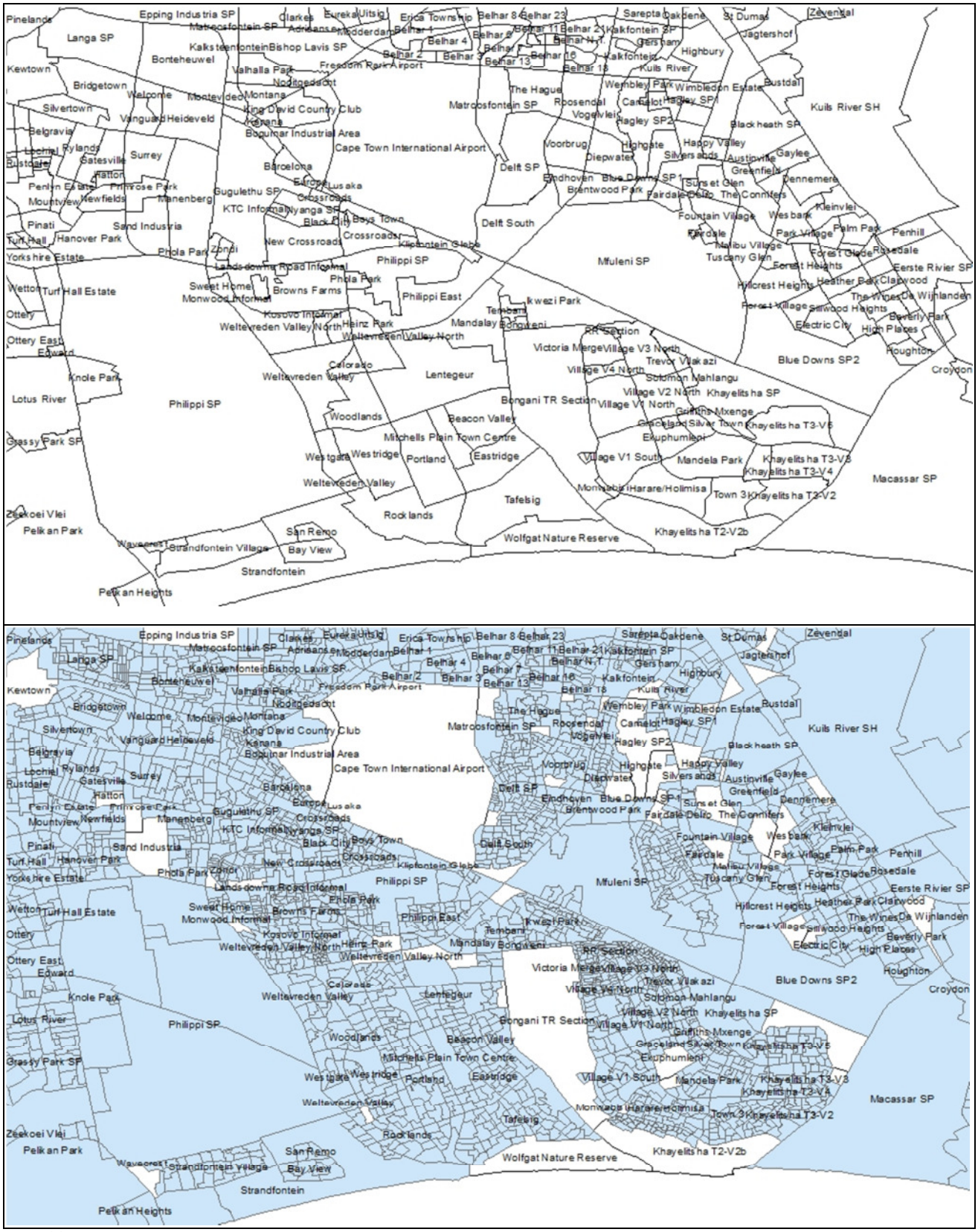
7 Appendix A

Table A1: Variables used to generate the wealth index using PCA

Category	Variables
Labour market status:	Employed, unemployed, discouraged work seeker, not economically active
Education:	No education, some primary education, secondary education, matric, higher/further education
Household income:	No income, low income (1 – 38200 rand) , middle income (38201 - 153800 rand), high income (153800 – 2457601 rand or more), Unspecified
Household assets:	Cell phone, computer, motor car, refrigerator, satellite tv, stove, tv, washing machine
Household services:	Water source –regional water scheme, Waste removal –local authority weekly, Internet access - from home; from cell phone; from work, from elsewhere; no access to internet, Toilet: flush toilet; flush toilet septic tank
Household size:	Household size 1 - Household size 10+

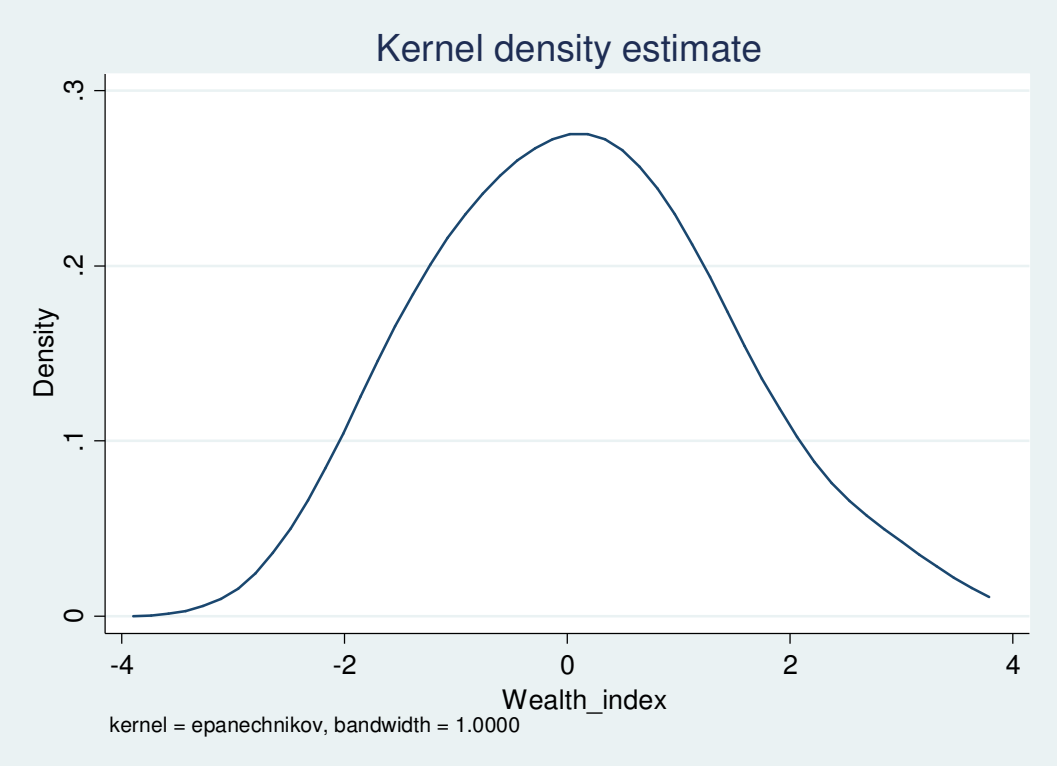
(Source: Census 2011)

Figure A1: Example of SP and SAL maps in metropolitan Cape Town



Note: The first map on top shows SP and the second map at the bottom the SAL layers. (Source: Census 2011 data)

Figure A2: Kdensity of PCA wealth index

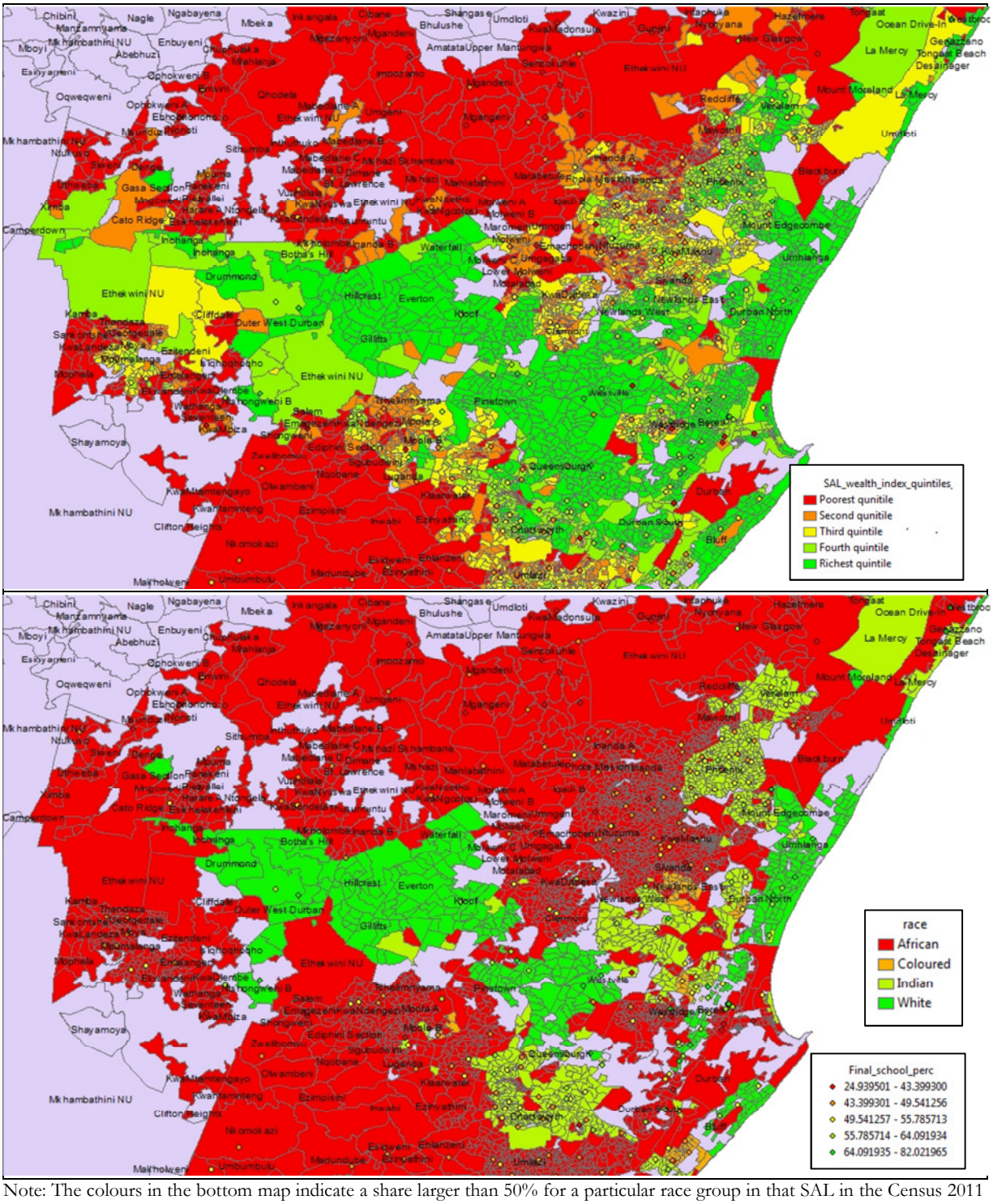


(Source: Census 2011)

Table A2: Racial distribution of age group 11-18 year olds for municipality of Cape Town (Source: Census 2011)

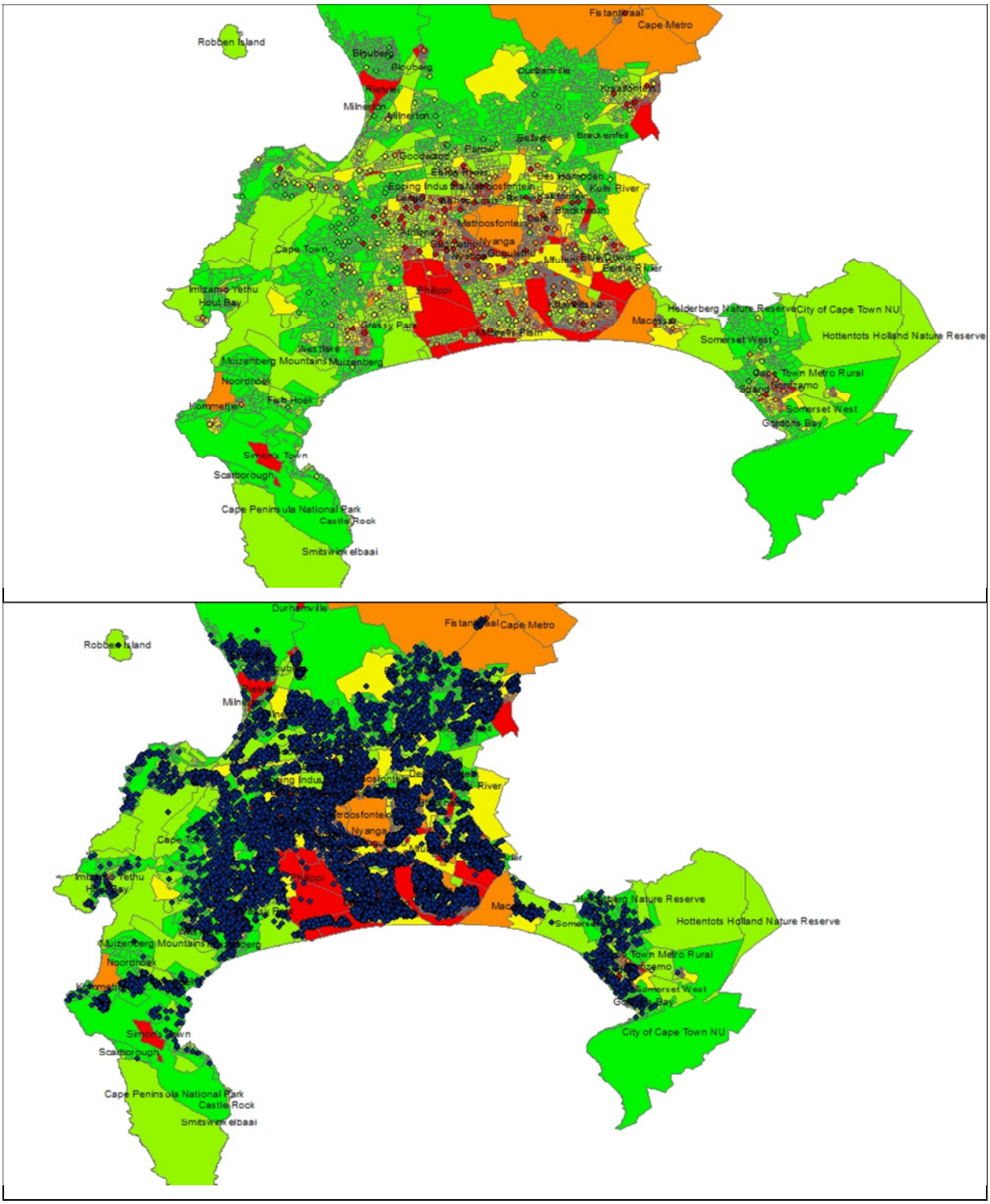
	Number	Percentage
Black	167786	38%
Coloured	219130	49%
Indian/Asian	5892	1%
Other	6457	1%
White	47597	11%
Total	446862	

Figure A3: Neighbourhood wealth quintiles, geographic race distribution and matric 2014 examination results in eThekwini metropolitan municipality (Source: Census 2011 and DBE school data).



Note: The colours in the bottom map indicate a share larger than 50% for a particular race group in that SAL in the Census 2011 data. In addition, the average matric 2014 examination results per school are displayed.

Figure A4: Map of Cape Town schools and CEMIS 2007 learner cohort (Source: master list of schools, CEMIS and Census 2011 data)



The map on top shows schools and the map at the bottom the learner distribution. The colour coding for neighbourhood wealth is green – richest, to red –poorest.

Figure A5: Example of Cape Town township schools and CEMIS 2007 learner cohort (Source: master list of schools, CEMIS and Census 2011 data)

The map on top shows schools and the map at the bottom the learner distribution. The colour coding for neighbourhood wealth is green – richest, to red –poorest.

Table A3 and Table A4 provide the average distance to school for each student in NIDS. The columns to the left provide the values for the new school distance variable which were obtained from the matching process of NIDS with the master lists of schools using fuzzy matching on school name and location. Given the significant differences to the match provided in the NIDS data (distance to school NIDS), in all the analysis of this study only the school information for wave 1-3 from the own matching was used.

Table A3: Distance to school

			Distance to school new			Distance to school from NIDS		
			km	std	Number	km	std	Number
Secondary age	Urban	Africa	18.68116	107.9719	3178	50.8819	158.7966	3191
		Coloured	10.43689	47.35696	863	91.61249	249.3532	822
		Indian	6.672051	6.829059	56	29.53672	163.9698	58
		White	27.06995	101.6134	139	77.66129	224.9168	14
	Rural	Africa	10.99625	36.3324	6898	28.64904	109.5429	6822
		Coloured	21.93374	45.39698	169	73.19331	227.7002	160
		Indian	7.298066	1.012356	24	25.892	89.19453	25
		White	44.2478	42.08362	9	396.2678	567.6691	9

(Source: NIDS wave 1-4)

Table A4: Distance to school

			Distance to school new			Distance to school from NIDS		
		Year	km	std	Number	km	std	Number
Primary age	Urban	2008	5.355357	23.17251	1837	39.87425	164.4093	1717
		2010	7.199118	45.95541	1775	45.70813	191.2405	1704
		2012	6.942235	35.83913	2239	25.70149	119.8447	1988
		2014				37.1221	153.8529	2739
	Rural	2008	9.102381	41.46133	2877	26.61582	116.4689	2681
		2010	7.182278	22.9229	2832	25.495	108.0987	2729
		2012	7.046202	25.88818	3536	20.7728	100.6671	3226
		2014				34.58738	238.4152	3817
Secondary age	Urban	2008	14.15244	68.49511	1305	58.58729	179.4829	1383
		2010	20.04269	137.484	1310	60.26372	181.1181	1275
		2012	17.14205	76.90393	1621	59.57468	188.3	1560
		2014				61.75101	179.7867	2014
	Rural	2008	11.82851	38.67698	2080	31.22783	120.053	2094
		2010	11.14124	37.08875	2281	29.84224	112.0515	2237
		2012	10.9952	34.43937	2739	29.50474	116.0339	2685
		2014				53.94867	492.6208	2952

(Source: NIDS wave 1-4)

Table A5: Difference in school and learner neighbourhood wealth quintiles

		Western Cape	Eastern Cape		
				11818	0.61%
	-3	5562	0.70%	80753	4.26%
School wealth quintile – student wealth quintile	-2	5537	0.70%	20864	1.10%
	-1	37597	4.74%	171369	9.03%
	0	427224	53.89%	1110119	58.50%
	1	178489	22.51%	199933	10.54%
	2	60356	7.60%	114254	6.02%
	3	78126	9.85%	171579	9.04%
				16789	0.88%

(Source: NIDS and Census 2011)

Table A6: Movement between school and student neighbourhoods for poorest 40%

		Frequency	Percentage
School wealth quintile – Student wealth quintile	-1	449739	6.20%
	0	3224914	47.50%
	1	1829967	27.00%
	2	492821	7.30%
	3	634828	9.40%
	4	158415	2.30%

This table uses the neighbourhood wealth quintile of the school and of the students' location from the Census 2011 community data (Source: Census 2011 and NIDS wave 1-3: age 14-18).

Table A7: OLS regression – Difference in school and learner neighbourhood wealth quintiles

VARIABLES	(1)	(2)	(3)
	Difference school and neighbourhood quintiles		
Male	-0.0352	-0.0247	-0.0274
	(0.0255)	(0.0255)	(0.0255)
Age	-0.00321	-0.00789	-0.00722
	(0.00938)	(0.00933)	(0.00932)
White	1.922***	1.907***	1.839***
	(0.532)	(0.531)	(0.531)
Coloured	0.975***	0.975***	0.942***
	(0.114)	(0.114)	(0.114)
Indian	3.238***	3.341***	3.283***
	(0.124)	(0.124)	(0.124)
Mother education	0.0174***		
	(0.00368)		
Father education	0.0236***		
	(0.00397)		
Ln(distance to closest school)	0.135***	0.135***	0.132***
	(0.0157)	(0.0157)	(0.0157)
Urban	1.073***	1.095***	1.080***
	(0.0363)	(0.0361)	(0.0362)
Father matric		0.308***	0.287***
		(0.0468)	(0.0470)
Mother matric		0.256***	0.223***
		(0.0431)	(0.0439)
Ln(per capita income)			0.0614***
			(0.0158)
Constant	0.177	0.451**	0.0789
	(0.193)	(0.189)	(0.211)
Observations	5 393	5 393	5 392
R-squared	0.348	0.348	0.350

Not reported province dummies (Source: NIDS wave 1-3)

Table A8: OLS regression – reaching and passing matric without repeating class (including distance) (Source: CEMIS)

VARIABLES	(1) Got matric	(2) Got matric	(3) Got matric	(4) Got matric
Coloured	0.0209***	0.0292***	-0.00421	-0.000890
	(0.00811)	(0.00581)	(0.00615)	(0.00618)
Indian	0.154***	0.136***	0.0475**	0.0397*
	(0.0230)	(0.0224)	(0.0225)	(0.0225)
White	0.144***	0.0939***	-0.0106	-0.0226*
	(0.0112)	(0.0119)	(0.0116)	(0.0127)
Overaged (in gr 6)	-0.280***	-0.283***	-0.273***	-0.272***
	(0.00550)	(0.00548)	(0.00545)	(0.00544)
Underaged (in gr 6)	0.0531***	0.0509***	0.0608***	0.0594***
	(0.0107)	(0.0107)	(0.0106)	(0.0106)
2. Learner neighbourhood quintile		0.0266***		0.0222***
		(0.00826)		(0.00822)
3. Learner neighbourhood quintile		0.0622***		0.0376***
		(0.00793)		(0.00854)
4. Learner neighbourhood quintile		0.182***		0.0852***
		(0.00809)		(0.00943)
5. Learner neighbourhood quintile		0.247***		0.0982***
		(0.0115)		(0.0129)
Ln(Distance to school)	0.00904***	0.0127***	-0.000513	-0.000182
	(0.00198)	(0.00192)	(0.00198)	(0.00198)
3. Old school quintile	-0.0472***			
	(0.0101)			
4. Old school quintile	-0.0286***			
	(0.0106)			
5. Old school quintile	0.130***			
	(0.0109)			
2. School average learner quintile			0.0211***	0.00546
			(0.00732)	(0.00782)
3. School average learner quintile			0.0711***	0.0421***
			(0.00800)	(0.00887)
4. School average learner quintile			0.161***	0.113***
			(0.00791)	(0.00946)
5. School average learner quintile			0.333***	0.270***
			(0.00946)	(0.0115)
Constant	0.381***	0.322***	0.335***	0.314***
	(0.00768)	(0.00802)	(0.00631)	(0.00819)
Observations	37 219	37 219	37 219	37 219
R-squared	0.141	0.143	0.157	0.159

Standard errors in parentheses, *** p<0.01, ** p<0.05, * p<0.1

Table A9: OLS regression – Standardized average matric results (Source: CEMIS and DBE school data)

VARIABLES	(1) Std (gr12) aggregate	(2) Std (gr12) aggregate
Coloured	-0.0415	-0.172***
	(0.0268)	(0.0255)
Indian	0.262***	0.205***
	(0.0696)	(0.0701)
White	0.353***	0.327***
	(0.0391)	(0.0388)
Age in years	-0.207***	-0.203***
	(0.0168)	(0.0170)
Ln(Distance to school)	0.0362***	0.0329***
	(0.00702)	(0.00708)
2.Learner neighbourhood quintile	0.0404	0.208***
	(0.0349)	(0.0361)
3. Learner neighbourhood quintile	0.0677**	0.371***
	(0.0332)	(0.0349)
4. Learner neighbourhood quintile	0.211***	0.761***
	(0.0337)	(0.0354)
5. Learner neighbourhood quintile	0.423***	1.249***
	(0.0406)	(0.0431)
2.School average learner quintile	-0.111***	
	(0.0360)	
3.School average learner quintile	-0.0258	
	(0.0373)	
4.School average learner quintile	0.212***	
	(0.0382)	
5.School average learner quintile	0.658***	
	(0.0421)	
Learner average - learner wealth quintile		0.198***
		(0.00956)
Constant	3.258***	3.053***
	(0.309)	(0.312)
Observations	12 131	12 131
R-squared	0.286	0.271

Standard errors in parentheses, *** p<0.01, ** p<0.05, * p<0.1

Table A10: Household characteristics for age 15-18

	Mean	std	Number
Per capita income	1520.6	2043.81	7245
Black	83.42%	0.31	7245
Coloured	8.48%	0.33	7245
Indian	2.30%	0.10	7245
White	5.80%	0.13	7245
Mother education	7.81	4.06	7245
Mother not in HH	0.16	0.38	7245
Father education	7.24	3.84	7245
Father not in HH	0.38	0.49	7245

Table uses population weights to obtain national representative results (Source: NIDS wave 1-3)

Table A11: Pooled OLS regression: reached years of education by age 15-18 with DBE quintiles

VARIABLES	(1) Years of education	(2) Years of education	(3) Years of education	(4) Years of education
Age	0.723***	0.709***	0.730***	0.726***
	(0.0222)	(0.0209)	(0.0207)	(0.0204)
Male		-0.513***	-0.517***	-0.521***
		(0.0805)	(0.0786)	(0.0792)
White		0.260	-0.130	-0.256
		(0.178)	(0.168)	(0.168)
Indian		0.873***	0.475**	0.357*
		(0.196)	(0.186)	(0.191)
Coloured		0.241*	0.174	0.116
		(0.127)	(0.118)	(0.125)
Mother education			0.0683***	0.0628***
			(0.00869)	(0.00875)
Mother not in the household			-0.165**	-0.178**
			(0.0793)	(0.0788)
Father education			0.0471***	0.0394***
			(0.00783)	(0.00783)
Father not in the household			-0.130**	-0.106**
			(0.0514)	(0.0499)
Ln(percy)				0.128***
				(0.0265)
2. DBE school quintile	0.119	0.0513	-0.0634	-0.0601
	(0.134)	(0.115)	(0.0994)	(0.0993)
3. DBE school quintile	0.251*	0.203*	0.0326	0.0114
	(0.130)	(0.118)	(0.101)	(0.102)
4. DBE school quintile	0.667***	0.470***	0.173	0.151
	(0.122)	(0.123)	(0.110)	(0.109)
5. DBE school quintile	1.020***	0.608***	0.114	0.0136
	(0.130)	(0.141)	(0.131)	(0.134)
Constant	-3.547***	-3.053***	-3.916***	-4.514***
	(0.364)	(0.374)	(0.368)	(0.401)
Observations	7 254	7 254	7 247	7 245
R-squared	0.263	0.311	0.356	0.359

(Source NIDS wave 1-3)

Table A12: Pooled OLS regression: reached years of education by age 15-18 with additional controls

VARIABLES	(1) Years of education	(2) Years of education	(3) Years of education
Age	0.705***	0.695***	0.696***
	(0.0223)	(0.0218)	(0.0218)
Male	-0.493***	-0.498***	-0.495***
	(0.0767)	(0.0773)	(0.0767)
White	-0.403**	-0.449**	-0.469**
	(0.177)	(0.178)	(0.182)
Indian	0.218	0.177	0.168
	(0.195)	(0.196)	(0.196)
Coloured	0.154	0.124	0.119
	(0.111)	(0.113)	(0.117)
Ln(percy)	0.107***	0.123***	0.120***
	(0.0318)	(0.0288)	(0.0290)
Mother education	0.0628***	0.0596***	0.0588***
	(0.00934)	(0.00886)	(0.00881)
Mother not in the household	-0.150**	-0.155**	-0.160**
	(0.0743)	(0.0747)	(0.0752)
Father education	0.0366***	0.0344***	0.0344***
	(0.00853)	(0.00774)	(0.00774)
Father not in the household	-0.110**	-0.101**	-0.0981**
	(0.0486)	(0.0497)	(0.0495)
Ln(School distance new)	0.0374*	0.0357	0.0330
	(0.0226)	(0.0226)	(0.0228)
Neighbourhood wealth quintiles	YES	YES	YES
2. School neighbourhood quintile	0.0810	0.103	0.106
	(0.113)	(0.112)	(0.113)
3. School neighbourhood quintile	0.158	0.182*	0.179
	(0.112)	(0.110)	(0.111)
4. School neighbourhood quintile	0.127	0.144	0.133
	(0.125)	(0.125)	(0.125)
5. School neighbourhood quintile	0.277*	0.275*	0.251*
	(0.141)	(0.141)	(0.138)
Learner-teacher ratio		0.920	0.568
		(2.224)	(2.121)
SGB teacher share		0.0391	
		(0.208)	
Private school			0.134
			(0.215)
Mixed funded school			0.0881
			(0.101)
Constant	-4.213***	-4.113***	-4.089***
	(0.396)	(0.408)	(0.409)
Observations	7 131	7 050	7 057
R-squared	0.354	0.356	0.356

(Source NIDS wave 1-3)