International Journal of Educational Development 33 (2013) 436–447

Contents lists available at SciVerse ScienceDirect

International Journal of Educational Development

journal homepage: www.elsevier.com/locate/ijedudev

Poverty & privilege: Primary school inequality in South Africa

Nicholas Spaull *

Stellenbosch University, 7600, South Africa

ARTICLE INFO

Keywords:
Primary schooling
South Africa
SACMEQ
Educational inequality
Student performance

ABSTRACT

The strong legacy of apartheid and the consequent correlation between education and wealth have meant that, generally speaking, poorer South African students perform worse academically. Although racial segregation has been abolished for 18 years now, schools which served predominantly White students under apartheid remain functional, while those which served Black students remain dysfunctional and unable to impart the necessary numeracy and literacy skills students should be acquiring by this level. The present study provides an overview of this dualistic nature of the primary education system in South Africa, with special attention paid to the bimodality of student performance. It argues that there are in fact two different education systems in South Africa and thus two different data-generating processes. These two sub-systems can be seen when splitting student performance by former-department, language, or socioeconomic status. The implications of such a dualistic schooling system are also elucidated, with special emphasis on government reporting and econometric modeling. The recently released SACMEQ III dataset is used for the econometric modeling. The study finds that when modeling student performance separately for the wealthiest 25% of schools on the one hand, and the poorest 75% of schools on the other, there are stark differences in the factors influencing student performance. Only five of the 27 factors are shared between the two models for mathematics, and 11 of the 30 factors for reading. This suggests a bifurcated system where the process which converts inputs into outputs is fundamentally different for each sub-system. Ultimately the paper has two logical conclusions: 1) Observing averages in South African education is uniquely misleading and overestimates the educational achievement of the majority of students, and 2) Modeling a single schooling system when there are in fact two school systems can lead to spurious results and misleading policy conclusions.

1. Introduction

In the years following the political transition in South Africa, the most important item on the national agenda was the social, economic and political integration of all South African people, particularly those marginalized under apartheid. After decades of systematic segregation and legislated racial exclusivity, the post-apartheid government faced the mammoth task of expanding service delivery, reducing widespread unemployment, and facilitating economic growth. As a means to this end, and to promote social cohesion, education was prioritized as an area for expansion and reform. Under apartheid there were multiple racially defined departments of education, each of which provided very different types and qualities of education based on the perceived role of that race-group in the apartheid society. Given the centrality of education to the inculcation and maintenance of the apartheid ideology, it is unsurprising that this area of social policy was highlighted for systemic reform in the post-apartheid years.[1]

However, while there was a sharp break in political ideology between the pre and post apartheid governments, many of the country's social institutions, such as schools, continued to function as they did under apartheid. The racially-defined departments were abolished in favor of nine provincial Departments of Education which operated in collaboration with a single national Department of Education. This being said, schools were and are still managed and run at the school-level by principals and, in the new dispensation, also by school governing bodies. Although the formal schooling institutions of apartheid were abolished (particularly racial segregation and inferior curricula), the informal schooling institutions inherent in non-White[2] schools remained largely intact. These ongoing informal institutions of disorder, distrust, rebellion, and lack of cooperation have undermined efforts to create an appropriate culture of teaching and learning in these schools. As the African National Congress (ANC) noted in 1994,

* Fax: +27 218084637.
E-mail addresses: spaull@sun.ac.za, nicholasspaull@gmail.com.

[1] The author would like to thank Servaas van der Berg and two anonymous reviewers for helpful comments on an earlier draft of this paper.

[2] The use of race as a form of classification and nomenclature in South Africa is still widespread in the academic literature with the four largest race groups being Black African, Indian, Colored (mixed-race) and White. This serves a functional (rather than normative) purpose and any other attempt to refer to these population groups would be cumbersome, impractical or inaccurate.

0738-0593/$ – see front matter
http://dx.doi.org/10.1016/j.ijedudev.2012.09.009

"Apartheid education and its aftermath of resistance has destroyed the culture of learning within large sections of our communities, leading in the worst-affected areas to a virtual breakdown of schooling and conditions of anarchy in relations between students, teachers, principals, and the education authorities" (ANC, 1994).

Fiske and Ladd (2004, p. 59) further elaborate on this concept and explain that in low income families there was a "lingering fear of education as an instrument of political subjugation" in the years following the political transition. Partly as a result of this social inertia, in combination with a host of other factors, many of the ex-Black schools which were entirely dysfunctional under apartheid remain largely dysfunctional today. They are characterized by severe underperformance, high grade repetition, high dropout, and high teacher absenteeism (Taylor et al., 2003; Fleisch, 2008). While many of these factors are certainly attributable to the socio-economic disadvantage of the students they serve, there is also an undeniable impact of more intangible elements such as ill discipline, inefficient management, and low cognitive demand – all legacies of apartheid. This low quality of education is further accentuated when compared to former 'Model-C' schools (ex-White) which are not dissimilar to schools in developed countries – both in terms of educational inputs and educational outcomes.

The aim of this paper is twofold: Firstly, it will provide an overview of the dualistic nature of the primary education system in South Africa, with special attention paid to the bimodality of student performance. Secondly, it will explore the implications of this bimodality by modeling the numeracy and literacy performance of South African Grade 6 primary school children. To do so, the recently released Southern and Eastern African Consortium for Monitoring Educational Quality (hereafter SACMEQ III) South African dataset is used. Following on from this, the implications for government reporting and modeling of South African student performance are also elucidated.

2. South Africa: a nation divided

Apartheid, which literally means 'separateness' in Afrikaans, aimed to create separate, and racially homogenous states, each of which would be ruled by its own people. While this aim was never fully realized, the systematic racial segregation practiced under apartheid, in conjunction with an overtly white supremacist ideology, has had, and continues to have a profound impact on the face of South African society. Eighteen years after the political transition, race remains the sharpest distinguishing factor between the haves and the have-nots, and while the upper-class of society is no longer entirely White – due largely to the slow emergence of a Black middle class – approximately 90% of the South African poor are Black (Leibbrandt et al., 2011). The links between affluence and educational quality in South Africa can partially explain this outcome since the poor receive a far inferior quality of education when compared to their wealthier counterparts (Van der Berg, 2007). It is now well established in the literature that the quality and duration of schooling that individuals receive is directly correlated with their labor-market prospects. Consequently, offering an inferior quality of education to the poor disadvantages them in the labor-market and entrenches their poverty. What is all the more disconcerting is that this does not refer to a minority of students, but rather the vast majority of the student population.

In the previous decade, South African primary education has been the subject of much research and debate. Of particular importance are the results of national and international assessments of student achievement. Internationally, South Africa has participated in three major cross-national comparisons of primary school student achievement, namely: SACMEQ[3] (2000 and 2007, Grade 6), TIMSS (2003, Grade 8), and PIRLS (2006, Grade 4 and 5), as well as a host of national standardized testing programs, the most important of which are the Systemic Evaluations (2001 and 2007, Grade 3), National School Effectiveness Study – NSES (2007–2009, Grades 3–5), and most recently, the Annual National Assessments – ANA (2011, Grades 1–6). All of these datasets have been analyzed by academic researchers, policy-makers and educational NGO's yielding a considerable amount of insight[4] into the performance of South African students, and the generative mechanisms behind that performance (Van der Berg et al., 2011; Carnoy et al., 2011). Unfortunately the picture that emerges time and again is both dire and consistent: However one measures learner performance, and at whichever grade one chooses to test, the vast majority of South African primary school learners are significantly below where they should be in terms of the curriculum, and more generally, have not reached a host of normal literacy and numeracy milestones.

National averages of 30–35% on tests of numeracy and literacy are the norm for tests calibrated to measure grade-appropriate performance as a 50% score, and can be seen in both the NSES (Taylor, 2011a) and the ANA evaluations (Department of Basic Education, 2011). Similarly, South Africa either has the lowest average score of all low income countries participating in international assessments (as in TIMSS and PIRLS), or, when the sample is limited to only Sub-Saharan Africa (as in SACMEQ) performs worse than many other countries which are considerably poorer, such as Kenya, Swaziland and Tanzania (Hungi et al., 2010).

The national averages reported above shroud the severe inequalities that plague all elements of South African life, and this is particularly true in education. It is now commonly accepted that when looking at learner performance in South Africa there is a minority of learners (roughly 25%) who attend mostly functional schools and perform acceptably on local and international tests while the majority of learners (roughly 75%) perform extremely poorly on these tests (Fleisch, 2008). Thus, there is a bimodal distribution of achievement in the country. As a result, the median reading score (SACMEQ – 464) is significantly lower than the mean (SACMEQ – 495), that is to say that the better performing 25% of students raise the extremely low average of the bottom 75%. Consequently, national averages overestimate the performance of the majority of South African learners since the distribution is skewed to the right. Because of this, the 'average' South African learner does not exist in any meaningful sense. However misleading this measure is, the national and provincial averages of learner performance remain the most commonly reported measure of achievement in government and international reports.

3. Bimodality

The bimodality of South African student performance is impervious to the grade or subject under assessment or the dataset under analysis. It can be seen as early as Grade 3 (Taylor, 2011a), and remains unabated until the national school leaving exam (Van der Berg, 2007). Furthermore, the bimodality of performance can be seen not only when the sample is split by wealth quartiles (Fig. 1 – SACMEQ), but also by school language (Fig. 2 – PIRLS), and former-department(Fig. 3 – NSES). This is unsurprising given the strong correlations between language,

[3] SACMEQ – Southern and Eastern African Consortium for Monitoring Educational Quality, TIMSS – Trends in International Mathematics and Science Study, PIRLS – Progress in International Reading and Literacy Study.

[4] The most comprehensive reports for each of these datasets are as follows: SACMEQ (Moloi and Chetty, 2011), TIMSS (Reddy, 2006), PIRLS (Howie et al., 2008), Systemic Evaluations (Department of Education, 2008), National School Effectiveness Study (Taylor, 2011b), and the Annual National Assessments (Department of Basic Education, 2011).

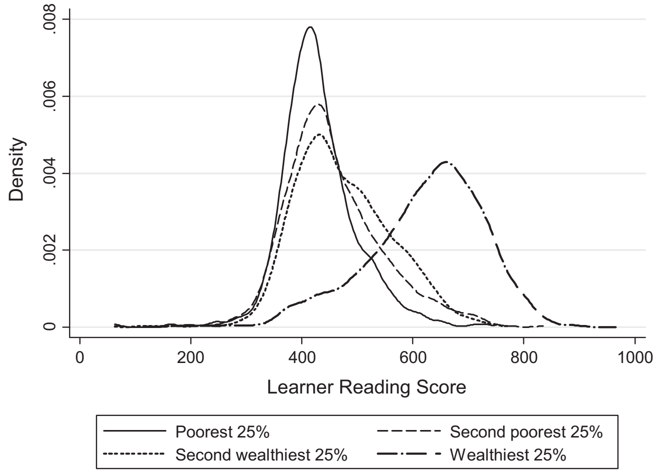

Fig. 1. Distribution of grade 6 reading performance by school wealth quartile (Data: SACMEQ III 2007). Source: Own calculations on SACMEQ III.

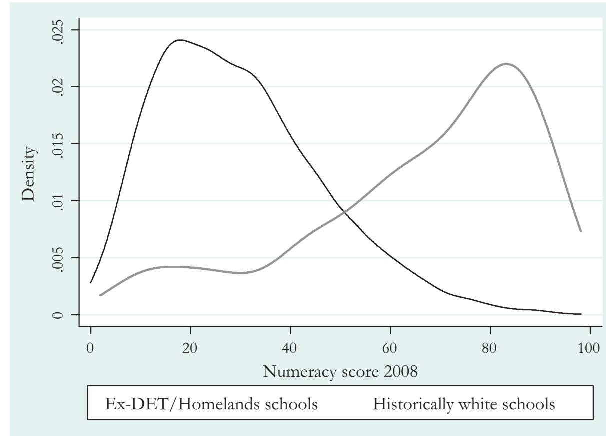

Fig. 3. Distribution of grade 4 numeracy achievement by historical education department (Data: NSES 2007/8/9). Source: Taylor (2011a).

socioeconomic status, and current school choice. The fact that these three figures are drawn from three independently conducted surveys at three different grades and at three different points in time further illustrates the consistency of the bimodal distribution in South Africa.

Importantly, it is not only student performance that is distributed in this dualistic way, but also various school level and home-background factors. Observing Table 1 shows that the wealthiest quartile (25%) of students seem to attend vastly differing schools than the remaining three quartiles (75%). In top quartile schools students are far more likely to have their own textbook, receive homework frequently, experience less teacher absenteeism, repeat fewer grades, live in urban areas, speak English more frequently at home, and have more educated parents (Table 1). All of these factors are likely to contribute to the better performance of this school sub-system. It is important to note that there is not a steady progression in any of these measures from quartiles one to three (as is the case in most other countries); the poorest three quartiles all have similar levels of grade repetition, teacher absenteeism, and textbook access.

The main explanation behind the bimodality of the schooling system in South Africa is twofold: (1) For whatever reason, historically disadvantaged schools remain dysfunctional and unable to produce student learning, while historically advantaged schools remain functional and able to impart cognitive skills; (2)

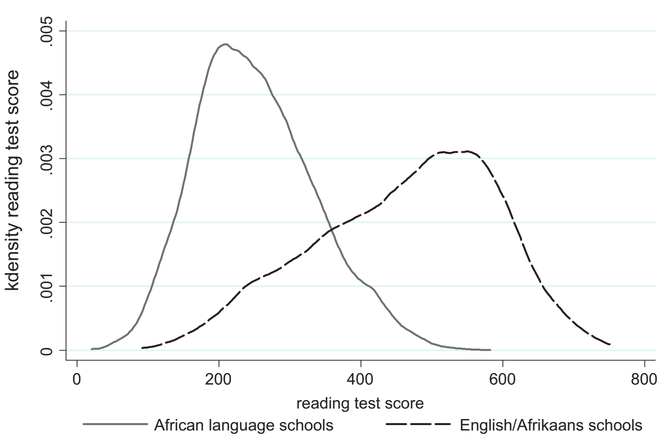

Fig. 2. Distribution of grade 5 literacy achievement by language of school (Data: PIRLS 2006). Source: Shepherd (2011)

The constituencies of these two school systems are vastly different with the historically Black schools still being racially homogenous (i.e. Black, despite the abolition of racial segregation) and largely poor; while the historically White and Indian schools serve a more racially diverse constituency, although almost all of these students are from middle and upper class backgrounds, irrespective of race. An important recent contribution by Yamauchi (2011) provides one explanation for this scenario. Using multiple data sources he shows that the spatial segregation policies of apartheid have had lasting impacts on the inequality of opportunity to quality education. Black students usually live far from good schools (situated in expensive neighborhoods), which make such schools geographically inaccessible, and those same schools usually charge higher school fees, which makes them financially inaccessible. Consequently, ex-Black schools have remained Black, while ex-White schools have become more racially diverse, albeit with wealthier Black, Colored and Indian students (Soudien, 2004).

The specific reasons for this bimodality are beyond the purview of this paper; they have been dealt with elsewhere in the literature (see Gustafsson, 2005; Fleisch, 2008; Van der Berg et al., 2011; Taylor, 2011a). It is sufficient for the purposes of this paper to accept that there is in fact a bimodal distribution of performance in South Africa, and that there are in fact two types of school systems, largely split along historical-school-system and socioeconomic lines. This is not a hypothesis, but rather stating one of the consistent characteristics of education data in South Africa.

Following on from the preceding discussion on the bimodality of the South African schooling system, it is worth asking whether the Department of Basic Education in South Africa has taken sufficient cognizance of this feature of the South African schooling system. Two areas where it has particular relevance are (1) for descriptive or reporting purposes, and (2) for prescriptive or policy-making purposes. These two areas are dealt with in the remaining part of this paper in sections four and five respectively.

4. Implications of a dualistic education system

4.1. Descriptive and reporting purposes

The practice of regularly reporting educational statistics is important for a variety of reasons. The most prominent of these is to tell if an education system is improving or deteriorating over time, which is necessary for accountability purposes, as well as being able to ascertain what does and does not work. This is especially true in developing countries where it is possible for relatively large changes in educational outcomes to occur in a

Table 1
Distribution of various schooling statistics across school wealth quartiles (Grade 6 – SACMEQ III).

Category	Variable	School wealth quartiles				Total	Quartiles relative to national average				Total
		1	2	3	4		1	2	3	4	
Performance	Reading score	430.5	457.8	474.0	623.7	494.9	−13%	−8%	−4%	26%	0%
	Mathematics score	450.9	467.1	470.7	593.8	494.8	−9%	−6%	−5%	20%	0%
	Proportion functionally illiterate[a]	43.3%	33.3%	25.6%	4.1%	27.3%	59%	22%	−6%	−85%	0%
	Proportion functionally innumerate	56.9%	48.6%	44.8%	8.4%	40.2%	42%	21%	12%	−79%	0%
	Reading teacher reading score	731.8	738.9	732.9	827.0	757.7	−3.4%	−2.5%	−3.3%	9.1%	0%
	Math's teacher mathematics score	719.6	729.1	751.7	863.5	763.6	−5.8%	−4.5%	−1.6%	13.1%	0%
Textbooks	Has own reading textbook	34.4%	42.3%	38.2%	66.1%	45.0%	−24%	−6%	−15%	47%	0%
	Has own mathematics textbook	27.6%	35.8%	32.3%	50.9%	36.4%	−24%	−2%	−11%	40%	0%
School factors	Gets homework "Most days of the week"	49.9%	52.1%	46.1%	75.8%	56.1%	−11%	−7%	−18%	35%	0%
	Self-reported teacher absenteeism (days)	24.2	22.7	20.1	11.6	19.7	23%	15%	2%	−41%	0%
	Repeated at least 2 grades	10.9%	9.3%	10.3%	1.8%	8.1%	34%	15%	27%	−78%	0%
	Pupil-Teacher-Ratio	36.3	34.8	35.5	30.5	34.3	6%	1%	3%	−11%	0%
	School in urban area	5.5%	21.4%	31.2%	73.3%	31.9%	−83%	−33%	−2%	130%	0%
	Student very old (14y+)	23.7%	20.1%	14.0%	2.0%	15.3%	55%	31%	−9%	−87%	0%
Home background	Speaks English at home 'Always'	5.6%	7.4%	9.2%	39.5%	15.3%	−64%	−52%	−40%	158%	0%
	Student has used a PC before	11.8%	39.9%	51.4%	94.9%	47.8%	−75%	−16%	7%	99%	0%
	More than 10 books at home	17.3%	23.0%	30.8%	67.2%	34.1%	−49%	−33%	−10%	97%	0%
	At least one parent has matric	29.9%	40.6%	49.3%	77.2%	48.5%	−38%	−16%	2%	59%	0%
	At least one parent has a degree	4.7%	7.8%	10.7%	28.7%	12.8%	−63%	−39%	−16%	125%	0%

Source: Own calculations based on SACMEQ III (2007) data.
[a] By this definition, a functionally illiterate learner cannot read a short and simple text and extract meaning, while a functionally innumerate learner cannot translate graphical information into fractions or interpret everyday units of measurement. See Shabalala (2005, p. 222) and Spaull (2011, p. 33) for further information.

single decade. In South Africa, most government reports present educational statistics at the provincial level of aggregation, rather than by wealth quartile or quintile[5] (see DBE, 2009, 2011 for examples). Given that the school system is administrated at the provincial level, and that provinces have a large degree of autonomy, this would seem to be the most logical practice. However, since there are two different types of education systems in South Africa distributed across all provinces, one functional, the other not, reporting mean achievement scores can potentially be very misleading. For example, national and provincial averages always overestimate the achievement of the majority of South African learners because the median is so far below the mean, as discussed in Section 2 above. Similarly for other measures of school functionality, averages shroud the true picture. Looking at self-reported teacher absenteeism, the national average of 19.7 days per year hides significant variation between the four wealth quartiles (Table 1). While teachers in quartile 1, 2 and 3 reported that they were absent for 24, 22, and 21 days respectively, teachers in the wealthiest quartile reported being absent for only 12 days in the preceding year. The same is true for functional illiteracy: while more than 25% of students in the poorest 3 quartiles are functionally illiterate, only 4% of quartile four students are thus classified (Table 1). If one only observed the national average, 27% functional illiteracy, this would not be apparent (see Spaull (2011b: 34) for a discussion on the definition of functional illiteracy used here).

Since there is reason to believe that schools, students and school sub-systems are far more homogenous within wealth quartiles than within provinces, it is somewhat perplexing that the former measure of aggregation is not used more frequently in government reports in addition to provincial averages, which are also important (although for different reasons). If student achievement was reported by wealth quartile in addition to province, one could determine if national trends are being driven by improvements in the wealthier or poorer schools, or one of a number of possible combinations of results. There is no reason to believe that these vastly differing school sub-systems would both rise or fall uniformly over time, as would be suggested if only a single average score was reported per province. If one agrees that there are indeed two underlying data-generating processes involved (and that average school wealth is strongly associated with performance), then reporting educational statistics by wealth quartile/quintile in addition to province is the most logical way forward.

4.2. Research purposes

The distinction between the two school sub-systems in South Africa is useful not only for descriptive purposes, but also for modeling and analytic purposes. In their report "Low quality education as a poverty trap" Van der Berg et al. (2011, 11) explain some of the quantitative reasons why modeling a single education system, when in fact there are two, can lead to spurious results:

> "There are important statistical and methodological reasons to analyze the two sub-systems separately when investigating what drives educational achievement in South Africa. Particular school inputs, teacher practices or other characteristics may affect student achievement differently across the two sub-systems. It is possible, for example, that an advanced media technology may be effective in the well-functioning system but ineffective in the historically disadvantaged system where schools may not have the expertise to implement the technology or the security to protect the equipment from theft and vandalism. In this way, important dynamics in one section of the school system can be glossed over by estimating a single model for the entire school system, alternatively, it is possible that a single model will suggest a relationship that is in fact invalid and is driven by differences between the two sub-systems. For example, it may be that within each sub-system additional resources do not produce improved student achievement, but that the one system has far superior resource

[5] Given that the distinction between quartile and quintile is small (5%), and that previous research and policy analysis has frequently used quintiles, the two are used interchangeably in this paper. Using kernel density curves, Spaull (2012) shows that the two parts of the SACMEQ III South African bimodal distribution are both most normally distributed when split by quartile rather than quintile.

endowments than the other and also produces better student achievement. Treating these two systems in a single model would suggest that additional resources do lead to better student achievement, when in fact this merely reflects overlapping differences between the two systems" (Van der Berg et al., 2011).

Given the bimodality of student performance seen in the SACMEQ III South Africa data (Fig. 1), and the dualistic nature of many educational inputs and indicators (Table 1), student performance is modeled in the next section using two separate models which estimates two separate data generating processes. The notion of a 'data-generating process' is of fundamental importance to statistical and econometric modeling. Underlying all modeling procedures is the assumption that there exists some real world data-generating process that has created the data found in a particular dataset – i.e. it is not entirely stochastic. In the context of the education production function, we assume that there is some knowable process by which educational outcomes (SACMEQ scores) are 'produced'. This process of production is assumed to be some function of a variety of 'inputs' such as parental education, educational resources, schooling environment etc. Since there is reason to believe that the South African schooling system is fundamentally bifurcated, it is also reasonable to suggest that there are two underlying data-generating processes and not one. Put differently, inputs are shaped and transformed into outputs in fundamentally different ways in the two South African schooling systems. In order to model these two separate data-generating processes, the regression analysis was done separately for each of the samples – 1) the wealthiest 25% of schools, and 2) the poorest 75% of schools in accordance with the bimodality shown in Fig. 1. If in fact there is a single data-generating process, the differences between the models for each sub-system should be minor.

4.2.1. Calculating school wealth

Before proceeding to the multivariate analysis, a more detailed description of the calculation of the school wealth variable is necessary. This is primarily because school wealth plays such a large role in the explanation of student performance in South Africa (Van der Berg, 2007; Yamauchi, 2011), but also because this is the variable on which the sample is split into wealthiest 25% of schools and poorest 75% of schools in this study.

Given the high degree of social and economic stratification in South Africa, school wealth can be accurately predicted based on the average socioeconomic status of the students in a school. Thus the question then becomes how one measures student socioeconomic status. Almost all methods of SES construction rely on some combination of wealth and parental education.[6] Due to the difficulties of extracting income data from children, most studies use asset-based questions as a proxy for household wealth (Filmer and Pritchett, 2001). Some authors have even posited that "asset-based measures of well-being may even be better than income or expenditure-based ones, since they may reflect the long-run welfare of the household better. They may also be more accurately measured (Filmer and Pritchett, 2001; Sahn and Stifel, 2003)" (cited in Wittenberg, 2009: 1). The researcher must then choose some method of aggregation ranging from a simplistic asset-count index, to the more complex measures of Principal Component Analysis (PCA) and Multiple Correspondence Analysis (MCA) which provide a linear combination of the asset variables which explains the greatest proportion of their joint variance. For the purposes of this study we used Multiple Correspondence Analysis due to the binary nature of the asset questions (Booysen et al., 2008), and used the 31[7] assets from the student questionnaire. Including parental education in the SES variable construction and as a separate variable in the model is obviously problematic since this introduces multicollinearity. As such, we did not include parental education in the SES variable because we wanted to include parental education as a separate variable. After calculating the student SES variable, the school SES variable was created as the average SES of all students in the school. Given that this new variable of school wealth is continuous, we used it to split the sample into the wealthiest 25% of schools and the poorest 75% of schools.

5. Modeling student achievement – SACMEQ III

In addition to identifying the characteristic features of the South African schooling system, it is revealing to model student achievement. Trying to understand the generative mechanisms of student learning can help educators and policy-makers alike, since it is through multivariate analysis that one can tease out which of the many possible explanatory variables are most closely associated with student learning. Given that many theoretically important variables are highly correlated with each other (especially in South Africa), it is often misleading to draw inference from bivariate distributions. For example, if one observes cross-tabulations between teacher knowledge and Grade 6 academic achievement in South Africa, there *seems* to be a clear relationship: more knowledgeable teachers teach better performing students. However, since better (and more knowledgeable) teachers select themselves into wealthier schools, and given the strong correlation between school wealth and a host of other factors in South Africa (socioeconomic status, first language, textbooks access and parental education) it is unclear whether the true generative mechanism here is more knowledgeable teachers, learning in a first language, or higher levels of parental education, or perhaps some combination of all of these factors or a variety of others. Bivariate analysis is unable to answer such questions. The unconditional correlation between two variables will often fall away when other variables are controlled for, indicating that some variables are merely proxying other important variables, and due to their correlation with those important variables there seems to be a strong relationship. Multivariate analysis can go some way to account for these correlations and provide some indication of the true generative mechanisms.

Using the SACMEQ III South Africa data we use an education production function approach to model the numeracy and literacy performance of Grade 6 South African students. While this approach has been used widely in the literature, there are also a number of statistical concerns when using this approach. These include omitted variable bias, sample selection bias, endogenous program placement and measurement error in the explanatory variables (Glewwe, 2002). While one can only truly deal with these issues by using panel data (see Rivkin et al., 2005), or randomized control trials (see Glewwe et al., 2012), by exercising caution in the interpretation of the results it is possible to draw useful inference from education production function analysis, even if one only

[6] Given that we are not really interested in school wealth in and of itself, but rather in its relation to the unmeasured elements of school quality and school functionality, one could also argue for a multidimensional approach which includes a variety of other correlates of school quality such as local policy indicators, efficiency indicators or social indicators. See Sengupta and Pal (2012) as one example.

[7] These 31 items were: daily newspaper, weekly or monthly magazine, clock, piped water, bore hole, table to write on, bed, private study area, bicycle, donkey/horse cart, car, motorcycle, tractor, electricity (mains, generator, solar), refrigerator/freezer, air-conditioner, electric fan, washing machine, vacuum cleaner, computer, internet, radio, TV, VCR player, DVD player, CD player, audio-cassette player, camera, digital camera, video camera, telephone/cell-phone (from Question 14 in Student Questionnaire).

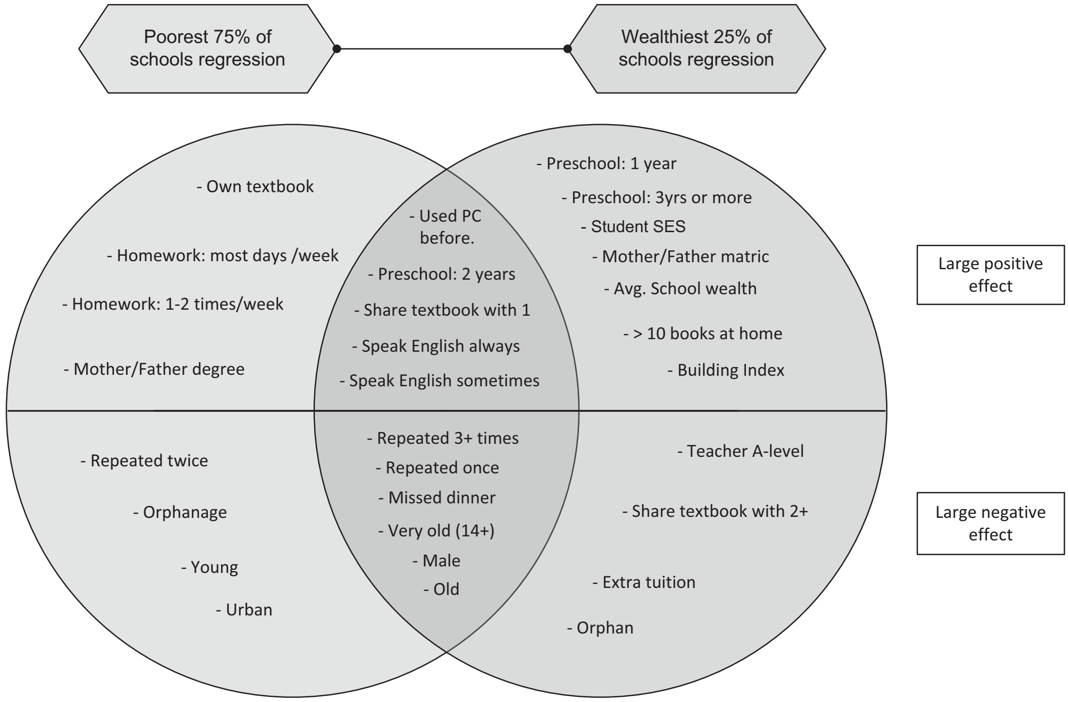

Fig. 4. Venn diagram for reading regressions (only large and significant coefficients shown).

interprets the coefficients as conditional correlations. Given the cross-sectional nature of the SACMEQ III dataset, alternative approaches wereunfortunately not possible in this instance.

5.1. Data

The Southern and Eastern African Consortium for Monitoring Educational Quality (SACMEQ) is a consortium of education ministries, policy-makers and researchers which, in conjunction with UNESCO's International Institute for Educational Planning (IIEP), aims to improve the research capacity and technical skills of educational planners in participating countries (Moloi and Strauss, 2005: 12). These surveys collect extensive background information on the schooling and home environments of students, and in addition, test students and teachers in both numeracy and literacy (see Ross et al. (2005) and Hungi et al. (2010) for a more in-depth discussion of the SACMEQ III numeracy and literacy tests). SACMEQ III South Africa tested 9071 Grade 6 students, and 1163 Grade 6 teachers from 392 schools. This nationally representative survey represents the most recent nationally representative international survey on educational quality in South Africa.

5.2. Method

To account for the complex two-stage survey design of SACMEQ (Ross et al., 2005), the statistical package STATA's built-in survey command was used in all regressions, with clustering by school and stratification by province – in accordance with the sampling structure used in the SACMEQ survey. More specifically, we used the Ordinary Least Squares (OLS) estimation procedure with Taylor linearization to approximate the variance of the point estimators. This accounts for the sampling structure, weighting and variance estimation in a complex two-stage survey design environment.

In each of the two models, 63 variables were used as explanatory factors in the determination of student numeracy and literacy performance. These variables were chosen based on pedagogical theory and included controls for province. Although the full list of 63 variables, and summary statistics for each variable, can be found in Appendix A, it is worth noting the major categories of variables: these include gender, age, student absenteeism, preschool education, home wealth, average student wealth of the school (average school wealth), missed meals, school location (urban/rural), parental education, English-use at home, orphan status, homework frequency, textbook access, grade repetition, class size, proportion of school's parents with matric and degree, teacher content knowledge, and teacher education.

The exact regression output from the two models can be found in the online appendix. If one observes only those coefficients that had a large impact on student performance (defined as a statistically significant coefficient that is greater than the absolute value of 0.1 standard deviations of student achievement) it becomes clear that that the factors that influence student performance are different for each of the two sub-samples (Figs. 4 and 5). The factors that had a large positive or negative impact on student reading performance can be found in Fig. 4, and for student mathematics performance in Fig. 5. This type of analysis is important since a variable may be statistically significant yet have little practical meaning in that its effect on student score is miniscule, although statistically different from zero.

6. Findings

6.1. Two data generating processes

The multivariate analysis of SACMEQ III, summarized in the Venn diagrams of Figs. 4 and 5, clearly supports the notion of two different school systems and thus two different data generating processes, in agreement with extant empirical studies on South African schooling (Van der Berg, 2008; Shepherd, 2011; Spaull, 2011a; Taylor, 2011a,b). Of the 30 variables that were large and significant in either of the reading models, only 11 are common to the two regressions (Fig. 4). That is to say that fewer than half of the variables that explain student performance in the wealthiest 25%

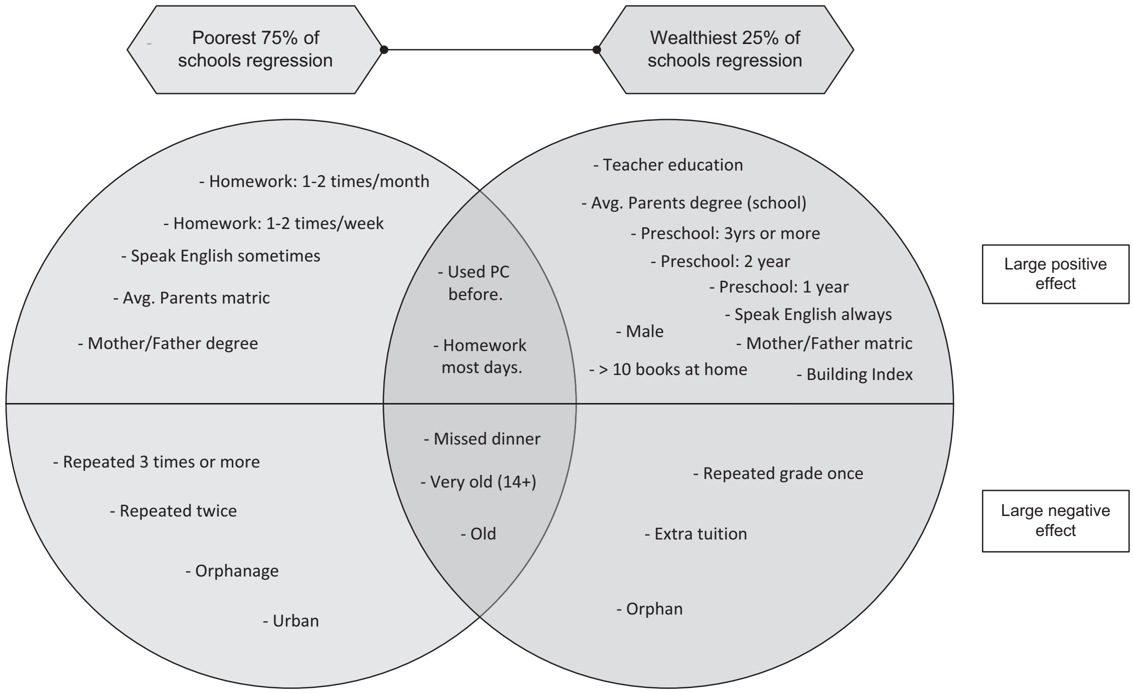

Fig. 5. Venn diagram for mathematics regressions (only large and significant coefficients shown).

of schools also explain student performance in the poorest 75% of schools. The situation is even clearer in the two mathematics regressions where only five of the 27 factors are common to the two models. Although many of the non-shared factors are similar in nature (for example repeated a grade once, twice, and three times in the mathematics regressions) it would be incorrect to assume that this was evidence of a similar data generating process. If one used a single model (i.e. full sample) for mathematics, this would force the coefficient on the dummy variable 'repeated a grade once' to be uniform across the two sub-systems, and similarly for 'repeated twice' and 'repeated three times or more' – in spite of the fact that these variables do not seem to be impacting student mathematics performance uniformly across the two sub-systems.

The impact of preschool education on Grade 6 performance provides a classic example of the conflation of effects that occurs when one mistakenly models a single data generating process when in fact there are two. If one uses a single model to predict the relationship between preschool education and Grade 6 literacy performance, there is a statistically significant and relatively large effect ranging from 0.09 standard deviations for one year of preschool to 0.15 standard deviations for two years of preschool education (see online appendix). However, this assumes an equal quality of preschool education across the socioeconomic spectrum. This is unlikely to be the case, since the quality of preschool education received by those of higher and lower socioeconomic status are likely to very different, as is well documented in the literature. For example, looking at developing countries in general Lee and Hayden (2009: 3) explain that "due to the proliferation of the variety of such [preschool] programsit has become critical to ensure program quality and administration". Similarly, Gustafsson (2010: 4) referring to the situation in South Africa states that "targeting publicly funded pre-primary services to the poor and ensuring that minimum quality criteria are upheld is particularly important if preprimary education is to have an impact". Indeed, the Presidency of South Africa in their "Medium Term Strategic Framework" has listed the improvement of the quality of early childhood education as a strategic priority (Presidency, 2009: 23). If one allows the impact of preschool education to differ for wealthy students and poorer students (by modeling the impacts separately), the difference in quality[8] becomes apparent. Relative to the combined (single) model, the impact of one year of preschool education is twice as large in the top quartile regression (0.17 standard deviations) as it is in the bottom three quartiles regression (0.06 standard deviations). Similarly, the impact of two years of preschool education is 230% larger for students attending the wealthiest quartile of schools (0.23 standard deviations) compared to students attending the poorest three quartiles of schools (0.1 standard deviations). This is to expected since students from wealthier backgrounds are more likely to attend preschools that are well equipped and staffed with trained professionals. In contrast, some preschool facilities offered to the poor are more accurately described as child-minding services, with unqualified staff, few educational resources, and little cognitive stimulation. Given these differences in the quality of preschool education in South Africa, it is unsurprising that the size of the impacts would differ across the two systems. Thus, modeling a single system overestimates the impact of low quality preschool education (offered to poorer students), and underestimates the impact of high quality preschool education (offered to wealthier students).

6.2. General findings

Of all the variables included in the models, there are four variables which are large and significant and shared by all four

[8] Since home socioeconomic status has already been controlled for, in addition to a variety of other factors which could influence early childhood education (parental education, pedagogical resources at home etc.), the impact of this preschool variable is likely to be a fairly accurate indicator of the impact of preschool education on student performance. Secondly, given that one year of preschool education is fairly common in both sub-systems with 63.3% of quartile1-3 students attending at least one year of preschool with the corresponding figure for quartile 4 students being 84.8 (see Appendix A), within each sub-system whether or not a student attends preschool is not likely to be proxying an innate parental value of education, which is another possible objection to this interpretation.

regressions (two reading and two mathematics regressions). Only one of the four (if a student has used a computer before) is positively associated with performance. Whether or not computer use is causally associated with reading and mathematics performance is unclear since this coefficient could simply be signaling an individual's preference for education, access to additional educational resources, or an unmeasured element of socio-economic status of the household or the school. The remaining three common variables found in all four regression models are all negatively associated with performance, and these are: whether student indicated that they normally missed dinner at least once per week, and whether a student was old or very old for Grade 6 (see Appendix B for variable descriptions). Clearly overage students do worse than their age-appropriate peers, even after accounting for the myriad of variables in the regressions. Interestingly the negative impact of 'missed dinner at least once per week' is the largest and most stable of the 'missed meals' type variables. This is to be expected since missing either breakfast or lunch once per week is a relatively common occurrence across wealth quartiles (see Appendix A: Summary statistics), while missing dinner once per week is less common. Thus, the missing dinner variable may be a better indicator of malnutrition or food poverty than the other 'missed meals' variables. Observing the R-squared values for the different regressions shows that the included variables are able to explain much more of the variation in student performance in the wealthiest quartile of schools, than in the poorest three quartiles.

When the sample is split into two groups, (1) the top quartile of school socioeconomic status and (2) the bottom-three quartiles of school SES, it is interesting to see that the included variables are able to explain more variation in wealthy learner performance (52% for reading and 44% for mathematics) than poorer learner performance (27% for reading and 17% for mathematics) (see Appendix C). The most likely cause of this difference is that variables that are important for understanding poor learners' performance have been excluded from the model. For example, variables such as school management and teacher quality are thought to be extremely important in understanding why some poor schools perform better than others. If the variation in school management and teacher quality is greater between poor schools than between wealthy schools, as we expect to be the case, then the exclusion of these variables will affect the bottom-three-quartile regression more than the top-quartile regression.

6.3. Factors associated with reading performance

The Venn diagram in Fig. 4 shows that there are significant reading benefits to speaking English at home either 'sometimes' or 'always'. Given that the SACMEQ tests were administered only in English and Afrikaans in South Africa,[9] one would expect there to be a benefit of speaking the language of the test at home. Sharing a textbook with only one learner was also associated with higher reading scores across both regressions, while sole use of a textbook was only large and significant in the poorer three quartile regression. Interestingly, higher homework frequency is only positively associated with reading performance in the poorer sample. One explanation for this is that 75.8% of students attending wealthy schools (top quartile) reported receiving homework most days of the week, compared to only 49.5% of students in the poorer sample (bottom three quartiles). Thus, the coefficients may be indicating that there are diminishing marginal returns to additional homework, or that poorer schools that prescribe homework more often are positively different to other poor schools in ways that are not captured by the other variables. Students attending poorer schools who had at least one parent with a degree did better than students where no parent had a degree. Only 7.4% of students from poor schools had at least one parent with a degree, compared to 28.7% in wealthier schools

In the wealthier subsample of schools, students that had preschool education did better than those that did not, as well as those with at least one parent who had completed matric. Additional resources at home ('more than ten books at home') and at school ('school building index') were associated with higher levels of student performance. The average wealth of students in the school was the largest single factor influencing student performance in wealthier schools.

In both models overage students did worse than age-appropriate students and boys did worse than girls in reading. Students who repeated a grade also did markedly worse than those that did not repeat. Looking only at the poorer schools' regression, the factors associated with lower performance were whether students were underage, whether they lived in an orphanage, and whether their school was in an urban area. By contrast, the factors associated with lower reading performance in wealthier schools were whether the child was an orphan, if the child attended extra tuition, or had to share a textbook with two or more students. The 'extra tuition' variable probably indicates which students were weaker to start with, and thus needed extra tuition, rather than a negative impact of additional teaching.

It is unclear why the teacher education dummy variable "Teacher A-level" is significantly associated with lower reading performance in wealthier schools, and further investigation is required.

6.4. Factors associated with mathematics performance

The only common factor in the mathematics regression which is not common in the reading regressions is the dummy variable "received homework most days of the week." Compared to students who never receive homework, any homework frequency is positively associated with mathematics performance in poorer schools. In addition, students who spoke English at home 'sometimes' did better than those who never spoke English at home.

In wealthier schools, the factors positively associated with mathematics performance are: teacher education, the average proportion of students in the school where at least one parent has a degree, if one parent has matric, home resources ('more than ten books at home'), school resources ('school building index'), speaking English at home 'always', and additional preschool education. Factors negatively associated with mathematics performance in wealthier schools are grade repetition (only one grade), extra tuition, and whether the student is an orphan. By contrast, higher levels of grade repetition (repeating a grade twice and three times or more) are negatively associated with mathematics performance in poorer schools, as is living in an orphanage or attending school in an urban area.

Three areas of the above analysis are worth closer inspection:

I. Comparing strongly associated factors across reading and mathematics performance *within* either the poorest 75% or wealthiest 25% of schools shows that many of the factors that affect literacy performance also affect numeracy performance.

[9] By Grade 6 almost all South African schools switch from home-language instruction to either English or Afrikaans instruction. This switch usually occurs at Grade 4.

II. Of all the factors considered, many are dummy variables with the same underlying question, for example, homework frequency, textbook access, frequency of grade repetition, duration of preschool, frequency of English spoken at home, and parental education. While the last three of these factors are not directly under the control of policy makers, grade repetition, preschool duration, and particularly textbook access (for reading) and homework frequency (for mathematics) are all areas where policy makers can exert some influence, indicating that there interventions that are likely to have a significant impact on student learning, and are also in the direct control of policy makers.

III. While the cut-off point of 0.1 standard deviations is somewhat arbitrary, any form of sub-classification requires that distinctions be made. For those readers who wish to verify these results, and check the specific point estimates for each coefficient, the full regression output (significant and insignificant coefficients with their respective significance levels) is presented in Appendix C, as well as the full South African model.

6.5. Robustness checks

Given the importance of socioeconomic status in South Africa and for the present analysis, as discussed in Section 4.2.1 above, we decided to run a series of robustness checks on the different models. As a model specification check we re-ran all three models (the combined sample, the wealthiest 25% of schools sample and the poorest 75% of schools sample) excluding the school SES variable and then excluding both the school SES and individual SES variables, i.e. in a step-wise fashion. The model is highly robust to these changes. In agreement with both the Frisch–Waugh–Lovell theorem and the consequences of omitted variable bias, excluding school socioeconomic status and individual SES leads to other variables that are correlated with these variables to become larger and/or significant. These variables include whether or not a student has used a computer before, the number of books at home, quality of school sanitation (in the poorer school sample), quality of school buildings (in the wealthier school sample), and school-level measures of the average parental education in the school. However, most of these variables are already larger than 0.1 standard deviations, or if they become significant after excluding the SES variables are smaller than 0.1 standard deviations. This means that the results reported in the Venn diagrams above remain almost exactly the same with or without the SES variables.

7. Conclusion

The motif that runs through much of the analysis above is that South Africa is still a tale of two schools: One which is functional, wealthy, and able to educate students; with the other being poor, dysfunctional, and unable to equip students with the necessary numeracy and literacy skills they should be acquiring in primary school. This dualistic nature of the education system has concrete implications for reporting educational statistics, and modeling educational performance. The South African government and external stakeholders should also report student performance by wealth quartile, and not only by province, principally because South African averages are uniquely misleading: they are not an accurate representation of any ‘average’ student, but rather overestimate the achievement of the majority of students. Secondly, since there are two data generating processes underlying these two systems, they should both be modeled separately. The Venn diagrams presented in the paper illustrated that there are very few variables that are common to both the top quartile of schools and the bottom three quartiles of schools when estimating factors associated with improved student performance. The example of preschool education illustrated this point well with estimates from a single combined model overestimating the impact of low quality preschool education and underestimating the impact of high quality preschool education, as was shown in the separate models. Consequently, modeling student performance in a single regression could lead to spurious results and misleading policy conclusions. If South Africa does indeed have two education sub-systems and not one, as has been argued in this paper, researchers and policy-makers would do well to take heed of this information when trying to understand educational data and formulate government policy. Without acknowledging and understanding the existing inequalities in South African primary education, particularly the extent and nature of those inequalities, the current patterns of poverty and privilege will remain unabated.

Appendix A. Summary statistics

See Table A1.

Table A1
Summary statistics.

Variable	Quartiles1-3 of school socioeconomic status		Quartile 4 of school socioeconomic status	
	Mean	Std. Dev.	Mean	Std. Dev.
Reading score	451.542	82.548	623.729	105.859
Mathematics score	461.616	72.980	593.821	96.558
Health score	475.782	87.037	583.217	97.588
R-teacher reading score	734.226	68.409	827.001	78.284
M-teacher maths score	731.349	84.588	863.503	114.784
Young (<11y3m)	0.037	0.188	0.028	0.165
Old (>11y3m–12y8m)	0.283	0.450	0.151	0.358
Very old (14y+)	0.197	0.398	0.020	0.141
Male	0.498	0.500	0.475	0.499
>5 Days absent	0.029	0.168	0.024	0.153
Preschool – months	0.053	0.224	0.035	0.185
Preschool – 1 year	0.358	0.479	0.253	0.435
Preschool – 2 years	0.134	0.341	0.213	0.409
Preschool – 3 years+	0.141	0.348	0.382	0.486

Table A1 (*Continued*)

Variable	Quartiles1-3 of school socioeconomic status		Quartile 4 of school socioeconomic status	
	Mean	Std. Dev.	Mean	Std. Dev.
SES	−0.406	0.804	1.005	0.787
SES squared	0.811	0.965	1.629	1.496
Lived with parents	0.688	0.463	0.854	0.353
3 or more siblings	0.630	0.483	0.279	0.449
Missed breakfast	0.287	0.453	0.274	0.446
Missed lunch	0.227	0.419	0.134	0.341
Missed dinner	0.149	0.356	0.062	0.241
>10 books at home	0.230	0.421	0.672	0.469
Used PC before	0.319	0.466	0.949	0.219
Urban	0.179	0.383	0.733	0.442
Mother/father matric	0.388	0.487	0.772	0.420
Mother/father degree	0.074	0.262	0.287	0.452
Spk Eng. at home sometimes	0.643	0.479	0.516	0.500
Spk Eng. at home always	0.072	0.258	0.395	0.489
Orphan (double-orphan)	0.104	0.305	0.050	0.219
Orphanage or children's home	0.008	0.087	0.004	0.065
School SES	−0.406	0.446	1.005	0.363
School SES squared	0.364	0.531	1.142	0.788
Homework – 1 or 2 times a month	0.102	0.303	0.040	0.196
Homework – 1 or 2 times a week	0.359	0.480	0.189	0.392
Homework – most days	0.495	0.500	0.758	0.428
Repeated a grade once	0.229	0.420	0.128	0.334
Repeated a grade twice	0.064	0.244	0.011	0.104
Repeated a grade three or more	0.039	0.193	0.007	0.082
No class library	0.619	0.486	0.360	0.480
Class-size > 40	0.630	0.483	0.319	0.466
Sanitation (Std)	0.454	1.026	−0.717	0.302
Building index (Std)	−0.480	0.771	1.157	0.626
Equipment index (Std)	−0.492	0.991	0.795	0.499
Avg. parent matric (school)	−0.385	0.733	1.173	0.703
Avg. parent degree (school)	−0.356	0.540	1.053	1.183
R-Textbook – teacher only	0.069	0.253	0.052	0.222
R-Textbook – share 2+	0.206	0.404	0.030	0.170
R-Textbook – share with 1	0.301	0.459	0.223	0.417
R-Textbook – own textbook	0.379	0.485	0.661	0.474
M-Textbook – teacher only	0.165	0.371	0.200	0.400
M-Textbook – share 2+	0.147	0.354	0.032	0.176
M-Textbook – share with 1	0.256	0.436	0.185	0.389
M-Textbook – own textbook	0.315	0.465	0.509	0.500
Extra Eng tuition	0.104	0.305	0.072	0.258
Extra Math tuition	0.093	0.290	0.114	0.318
R-Teacher male	0.329	0.470	0.227	0.419
R-Teacher Jnr. Secondary	0.019	0.138	0.011	0.103
R-Teacher Snr. Secondary	0.133	0.339	0.095	0.293
R-Teacher A-level/further study	0.166	0.372	0.119	0.324
R-Teacher degree	0.413	0.492	0.589	0.492
R-Teacher training: 2 yrs	0.088	0.284	0.067	0.251
R-Teacher training: 3 yrs	0.518	0.500	0.198	0.399
R-Teacher training: >3 yrs	0.366	0.482	0.640	0.480
M-Teacher male	0.428	0.495	0.268	0.443
M-Teacher Jnr. Secondary	0.022	0.148	0.001	0.032
M-Teacher Snr. Secondary	0.111	0.315	0.032	0.176
M-Teacher A-level/further study	0.157	0.364	0.199	0.399
M-Teacher degree	0.448	0.497	0.618	0.486
M-Teacher training: 2 yrs	0.073	0.260	0.016	0.124
M-Teacher training: 3 yrs	0.477	0.500	0.175	0.380
M-Teacher training: >3 yrs	0.433	0.496	0.728	0.445
Eastern Cape	0.213	0.410	0.013	0.114
Free State	0.054	0.225	0.037	0.190
Gauteng	0.109	0.311	0.361	0.480
KwaZulu Natal	0.248	0.432	0.204	0.403
Limpopo	0.159	0.366	0.060	0.237
Mpumalanga	0.097	0.297	0.050	0.218
Northern Cape	0.022	0.146	0.016	0.126
Western Cape	0.039	0.193	0.195	0.396

Appendix B. Variable descriptions

Variable	Description
Young (<11y3m) Old (>11y3m–12y8m) Very old (14y+)	According to the 2002 amendments of the South African Schools Act (2002, Section 5) a child may be admitted to Grade 1 if he or she is five turning six by 30 June in the year of admission, or he or she must wait until the following year to be admitted. Consequently, children can be deemed age-appropriate if they are between five years six months and seven years old at the beginning of Grade 1. Thus in Grade 6, the age range would be between 10 years six months and 12 years old. Given that the SACMEQ survey was administered in September, the age range for learners writing the SACMEQ tests would be between 11 years and three months and 12 years and nine months. The age categories were calculated as young (less than 11 years 3 months), age appropriate (11 years 3 months – 12 years 8 months), old (12 years 8 months – 13 years), and very old (14 years and older).
Male	Gender of the student
>5 Days absent	Self-reported student absenteeism
Preschool – months Preschool – 1 year Preschool – 2 years Preschool – 3 years or more	Students were asked "How long did you attend a preschool, kindergarten, nursery, reception, etc., before Grade 1? (Base category "no preschool")
SES SES squared	In SACMEQ III, as is the case with most surveys which target children, it is not possible to get an accurate representation of the monetary value of family income. Consequently, socioeconomic status (SES) was inferred from a series of possession questions. In SACMEQ III, learners were asked whether or not each of 31[1] items was found in the place where they stayed during the school week. To construct the SES variable, all 31 of these items were used in a Multiple Correspondence Analysis (MCA) forming the SES index. The SES variable was transformed to be the negative of the MCA index to ensure that the largest positive value of MCA was assigned to the wealthiest learner for ease of interpretation.
Lived with parents	Whether a student lived with their parents
3 or more siblings	Whether a student had 3 or more siblings
Missed breakfast	Whether student indicated that they normally missed breakfast at least once per week
Missed lunch	Whether student indicated that they normally missed lunch at least once per week
Missed dinner	Whether student indicated that they normally missed dinner at least once per week
More than 10 books at home	If a student indicated that they had more than ten books at home
Used PC before	If a student indicated that they had used a computer before
Urban	If the school a student attended was located in an urban area (base category = rural area)
Mother or father has matric	If a student indicated that either their father or their mother (or both) had a matric qualification
Mother or father has degree	If a student indicated that either their father or their mother (or both) had a degree
Speak Eng. at home sometimes	If a student indicated that they spoke English at home "sometimes" (base category is "never")
Speak Eng. at home always	If a student indicated that they spoke English at home "always" (base category is "never")
Orphan (double-orphan)	If a student indicated that both parents were deceased
Orphanage or children's home	If a student indicated that they stayed in an "orphanage or children's home" during the school week
School SES	The average SES of students in that school
School SES squared	School SES squared
Homework – 1 or 2 times a month Homework – 1 or 2 times a week Homework – most days	Homework frequency as indicated by the student – the question did not differentiate between reading homework or mathematics homework
Repeated a grade once Repeated a grade twice Repeated a grade three or more	Grade repetition as indicated by the student
No class library	Dummy variable indicating that the classroom did not have a library
Class-size >40	Class size as reported by the principal
Sanitation (Std)	A standardized measure of sanitation quality with the underlying variable being the number of non-flushing school toilets (latrine places, squat holes, or pit toilets) as a percentage of total toilets (non-flushing and flushing toilets combined). The higher the index the poorer the sanitation quality.
Building Index (Std)	Standardized measure of school buildings based on the underlying variable of the presence/absence of seven school buildings: school library, school or community hall, teacher/staff room, separate office for School Head, store room, special area for guidance and counseling, and cafeteria/shop/kiosk.
Equipment Index (Std)	Standardized measure of school buildings based on the underlying variable of the presence/absence of 18 items: first aid kit, clock, telephone, typewriter, duplicator, electricity (mains or generator), radio, tape recorder, TV, audio cassette player, CD player, VCR machine, DVD player, fax machine, photocopier, overhead projector, computer(s), computer room.
Avg. Parent matric (school)	A standardized measure of the proportion of students in a school where at least one parent had matric
Avg. Parent degree (school)	A standardized measure of the proportion of students in a school where at least one parent had a degree

Appendix B (*Continued*)

Variable	Description
R/M Textbook – teacher only R/M Textbook – share 2+ R/M Textbook – share with 1 R/M Textbook – own textbook	Accesibility of reading and mathematics textbooks as answered by the students. For the reading regressions the results from the reading textbook question were used, and for the mathematics regressions the results from the mathematics textbook question were used. in both instances the base category is "no textbook"
Extra tuition (English/Maths)	If a student indicated that they had attended extra tuition for reading (included in the reading regression) or mathematics (included in the mathematics regression)
Reading-teacher reading score	Reading teacher score scaled to be comparable to the student scores included n reading regression) – see Ross et al. (2005) for a full discussion
Mathematics teacher mathematics score	Mathematics teacher score scaled to be comparable to the student scores (included in mathematics regression) – see Ross et al. (2005) for a full discussion
Teacher Male	Gender of the teacher
Teacher Jnr. Secondary Teacher Snr. Secondary Teacher A-level/further study Teacher degree	Highest level of teacher education – answered by the reading teacher and the mathematics teacher.
Teacher training: >3 yrs	Highest level of teacher training – answered by the reading teacher and the mathematics teacher. Dummy variable equal to one if the teacher has more than three years of teacher training.

Appendix C. Supplementary data

Supplementary data associated with this article can be found, in the online version, at http://dx.doi.org/10.1016/j.ijedudev.2012.09.009.

References

ANC, 1994. A Policy Framework for Education and Training. African National Congress, Pretoria.

Booysen, F., van der Berg, S., Burger, R., von Maltitz, M., du Rand, G., 2008. Using an asset index to assess trends in poverty in seven sub-Saharan African countries. World Development 36 (6), 1113–1130.

Carnoy, M., Chisholm, L., Addy, N., Arends, F., Baloyi, H., Irving, M., et al., 2011. The Process of Learning in South Africa: The Quality of Mathematics Teaching in North West Province. HSRC Press.

DBE, 2009. Trends in Education Macro Indicators Report. Department of Basic Education, Pretoria.

DBE. (2011). Macro Indicator Trends in Schooling: Summary report 2011. Retrieved from Department of Basic Education: http://www.education.gov.za/LinkClick.aspx?fileticket=lkzSZYKJGno%3D& tabid=422&mid=1261.

Department of Basic Education, 2011. Report on the Annual National Assessments of 2011. Department of Basic Education, Pretoria.

Department of Education, 2008. Grade 3 Systemic Evaluation 2007 Leaflet. Department of Education, Pretoria.

Filmer, D., Pritchett, L., 2001. Estimating wealth effects without expenditure data-or tears: an application to educational enrollments in states of India. Demography 38 (1), 115–132.

Fiske, E., Ladd, H., 2004. Elusive Equity: Education Reform in Post-apartheid South Africa. Brookings Institution Press/HSRC Press, Washington.

Fleisch, B., 2008. Primary Education in Crisis: Why South African Schoolchildren Underachieve in Reading and Mathematics. Juta & Co, Cape Town.

Glewwe, P., 2002. Schools and skills in developing countries: education policies and socioeconomic outcomes. Journal of Economic Literature 40 (2), 436–482.

Glewwe, P., Hanushek, E., Humpage, S, Ravina, R., 2012. School Resources and educational outcomes in developing countries: A review of the literature from 1990 to 2010, Working Papers 120033, University of Minnesota, Center for International Food and Agricultural Policy.

Gustafsson, M., 2005. The Relationship Between Schooling Input and Output in SA – Methodologies and Policy Recommendations Based on 2000 SACMEQ. SACMEQ, Paris.

Gustafsson, M., 2010. Recent patterns in the extent and quality of pre-primary education in South Africa. Stellenbosch Economic Working Papers.

Howie, S., Venter, E., Van Staden, S., Zimmerman, L., Long, C., Du Toit, C., et al., 2008. PIRLS 2006 Summary Report: South African Children's Reading Literacy Achievement. Center for Evaluation and Assessment, Pretoria.

Hungi, N., Makuwa, D., Ross, K., Saito, M., Dolata, S., van Capelle, F., et al., 2010. SACMEQ III Project Results: Pupil Achievement Levels in Reading and Mathematics. Southern and Eastern Africa Consortium for Monitoring Educational Quality, Paris.

Lee, Y., Hayden, M., 2009. Early childhood care and education: worldwide challenges and progresses. Current Issues in Comparative Education 11, 2009.

Leibbrandt, M., Wegner, E., Finn, A., 2011. The policies for reducing income inequality and poverty in South Africa. South Africa Labour and Development Research unit (SALDRU). Working Paper Number 64.

Moloi, M., Chetty, M., 2011. SACMEQ The SACMEQ III study in South Africa: A Study of the Conditions of Schooling and the Quality of Education. Southern and Eastern Africa Consortium for Monitoring Educational Quality, Pretoria.

Moloi, M., Strauss, J., 2005. The SACMEQ II project in South Africa: A study of the conditions of schooling and the quality of education – South Africa working report. SACMEQ. Online. Available: http://www.sacmeq.org/education-south-africa.htm [September 2010].

Presidency, 2009. Medium Term Strategic Framework. The Presidency of the Republic of South Africa, Pretoria.

Reddy, V., 2006. Mathematics and Science Achievement at South African Schools in TIMSS 2003. HSRC Press, Cape Town.

Rivkin, S., Hanushek, E., Kain, J., 2005. Teachers, schools, and academic achievement. Econometrica 73 (2), 417–458.

Ross, K.N., et al., 2005. The conduct of the SACMEQ III project. In: Onsomu, E., Nzomo, J., Obiero, C. (Eds.), The SACMEQ II Project in Kenya: A Study of the Conditions of Schooling and the Quality of Education, SACMEQ, Harare.

Sahn, D., Stifel, D., 2003. Urban–Rural inequality in living standards in Africa, Journal of African Economies. Centre for the Study of African Economies (CSAE) 12 (4), 564–597.

Sengupta, A., Pal, N., 2012. Assessing the primary schools – a multi-dimensional approach: a school level analysis based on Indian data. International Journal of Educational Development 32, 264–272.

Shabalala, J., 2005. The SACMEQ II Project in Swaziland: A study of the conditions of schooling and the quality of education. Retrieved 2009, from Southern and Eastern Africa Consortium for Monitoring Educational Quality: http://www.sacmeq.org/education-swaziland.htm.

Shepherd, D., 2011. Constraints to school effectiveness: what prevents poor schools from delivering results? Stellenbosch Economic Working Papers.

Soudien, C., 2004. Constituting the class: an analysis of the process of 'integration' in South African schools. In: Chisholm, L. (Ed.), Changing Class: Education and Social Change in Post-Apartheid South African Schools. HSRC Press, Cape Town.

Spaull, N., 2011a. A preliminary analysis of SACMEQ III South Africa. Stellenbosch Economic Working Papers.

Spaull, N., 2011b. Primary School Performance in Botswana, Mozambique, Namibia and South Africa. Paris: Southern and Eastern African Consortium for Monitoring Educational Quality (SACMEQ) Working Paper no. 8.

Spaull, N., 2012. Equity & Efficiency in South African Primary Schools. Masters Thesis. Stellenbosch University.

Taylor, S., 2011a. Uncovering indicators of effective school management in South Africa using the national school effectiveness study. Stellenbosch Economic Working Papers.

Taylor, N., 2011b. National School Effectiveness Study – Synthesis Report. JET Education Services, Johannesburg.

Taylor, N., Muller, J., Vinjevold, P., 2003. Getting schools working. Pearson Education, Cape Town.

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. Centre for European, Governance and Economic Development Research (CEGE) Discussion Papers 69.

Van der Berg, S., Burger, C., Burger, R., de Vos, M., du Rand, G., Gustafsson, M., et al., 2011. Low Quality Education as a Poverty Trap. Stellenbosch University, Stellenbosch.

Wittenberg, M., 2009. Weighing the value of asset proxies: the case of the body mass index in South Africa. A Southern Africa Labour and Development Research Unit Working Paper Number 39. Cape Town: SALDRU, University of Cape Town.

Yamauchi, F., 2011. School quality, clustering and government subsidy in post-aparteid South Africa. Economics of Education Review 146–156.