UNRAVELLING THE MYSTERY: UNDERSTANDING SOUTH AFRICAN SCHOOLING OUTCOMES IN REGIONAL CONTEXT

Servaas van der Berg & Megan Louw
University of Stellenbosch
(email: SvdB@sun.ac.za; MeganL@sun.ac.za)

Paper to the conference of the Centre for the Study of African Economies, Oxford University, 21st March 2006

Introduction

A recent survey of reading and mathematics achievement of grade 6 students of fourteen countries[1] in Southern and Eastern Africa has provided researchers with a wealth of data with which to answer important questions about schooling outcomes in these countries, especially regarding school quality. One of the most intriguing issues to emerge from the second SACMEQ (Southern Africa Consortium for Monitoring Educational Quality) study, conducted in 2000, is alluded to in recent work by Lee, Zuze and Ross (2005). Adopting a multilevel approach that partitions variance in student achievement into a within-school component and a between-school component, the authors find that a striking 70 percent of the variation in South African reading performance occurs between schools. This can be contrasted with 20-60 percent for the remaining 13 countries included in the study, with between-school variation in Namibian achievement coming a rather distant second to South Africa's, at 59 percent (Lee et al. 2005: 16).

Further, it is interesting to note that South Africa's performance on the tests is lacklustre in regional context, given that students from this country benefit from having the third highest average socioeconomic status within the group, lagging only their counterparts from the island economies of Mauritius and Seychelles. In light of the significant role typically assigned to a student's socioeconomic status in the education production function literature, South Africa's position in the literacy test score ranks – 8th out of 14 – is surprising. Using the same SACMEQ dataset, Moloi (2005) from the South African national Department of Education found that more than half of grade 6 pupils in South Africa perform at a grade 3 level or lower in mathematics. Similarly, a member of the provincial Western Cape Education Department (Renault-Smith 2006: 230, 234, 238) admits that many schools had been dysfunctional, and that interventions to improve this situation have brought only minor improvements at the matriculation (school-leaving or grade 12) level.

The purpose of this paper is thus to unravel the mystery behind the numbers – to discover how and why educational outcomes are determined systematically differently in South Africa from the way they are in the rest of the region. As Lee et al. (2005) do, we also utilize data from the SACMEQ II study, although our concern lies with mathematics rather than literacy outcomes. Other international studies[2] have shown that South African students perform particularly poorly

---

[1] For these purposes, Zanzibar and Tanzania were treated as if separate countries.

[2] In the 2003 TIMMS (Trends in International Mathematics and Science Study), the scores of both former black and former white schools of 227 and 456 respectively on the Mathematics part of this grade 8 test placed both these groups of schools below the international average (467); the South African country average of 264 was very far below this international average. In fact, South Africa's score was the lowest of all 53 participating countries; even the other five much poorer African countries in the study outscored South Africa. On the 1995 MLA (Monitoring Learning Achievement) study, South African grade 4 pupils scored by far worst in numeracy of twelve participating African countries, whilst South Africa outperformed only 3 of the 12 countries in literacy. (Taylor, Muller & Vinjevold 2003: 41)

on numeracy tests, and it is our objective in the current paper to work towards an understanding of the factors underlying this finding.

Following a brief outline of the principles of education production function theory, we present descriptive statistics and the results of our preliminary investigation into explaining the apparent anomaly. These confirm our hypothesis that South African households and schools are richly endowed with resources in relative terms, suggesting that the inputs into this country's educational process are favourable. To identify the drivers of the observed mediocre educational outcomes in South Africa, we then turn to regression analysis, fitting aggregate models as well as separate ones for South Africa versus the countries comprising the remainder of the region. Both survey regressions and hierarchical linear models are estimated, and the results of analysis are presented and discussed. Delving into deeper issues concerning the generation of schooling outcomes in South Africa, we briefly review the work of Van der Berg (2005) and Gustafsson (2005 & 2006). The paper concludes with a summary of our findings and by highlighting the policy relevance of these.

A brief overview of the education production function literature

Education production function theory is concerned with identifying which student inputs are important for determining individual schooling outputs, modelling student performance as the outcome of a technological process. Apart from the impact of a student's innate ability, theory predicts that his/her schooling achievement is influenced by socioeconomic status, parents' education, school quality and a range of other variables, including peer- and community-level factors. Greater affluence implies that more household resources are available to support learning through purchasing textbooks and stationery, paying for additional tuition, and providing the student with sufficient nutrition to maximize learning. It may also raise the returns to a child's education through the beneficial impact of social networks on his/her future job prospects. Having more educated parents is expected to positively influence student achievement through a host of factors, including help with his/her homework, greater support for his/her education (implying that more of a given household budget is allocated towards schooling) and maintaining pressure on the management of the school at which he/she is enrolled to provide superior quality education.

Despite the strong intuitive appeal of the proposition that school quality matters for schooling outputs, there has been great division in the empirical evidence for this hypothesis. Hanushek (1986; 1995) is one of the most outspoken critics of the merit of employing education production studies to investigate this issue, although he has softened his stance somewhat in more recent years. However, there is a growing body of research confirming that school quality does influence student achievement in measurable ways (see for instance Krueger 1999; Angrist & Lavy 1999; Case & Deaton 1999). Fuller (1985) and Hanushek (1995) also find tentative evidence that school resource inputs are more strongly linked to schooling outputs in developing countries; textbooks, student's writing materials, teacher's tertiary education, and school facilities including libraries are highlighted as some of the most important determinants of student achievements in these nations (Fuller 1985; Hanushek 1995). One reason advanced for the apparent divergence across the industrialised and developing worlds is the existence of threshold effects operating with respect to school resource inputs (Fuller 1985). However, this finding should be accepted with great caution, since Fuller and Hanushek's work comprise literature surveys of a large number of studies that use datasets and regression methodology of widely varying reliability (Glewwe 2002).

Data and methodology

The 2000 SACMEQ II dataset comprises literacy and numeracy data for some 42 000 pupils from 2 300 "mainstream" government and non-government schools in Botswana, Kenya, Lesotho, Malawi, Mauritius, Mozambique, Namibia, Seychelles, South Africa, Swaziland, Tanzania and Zanzibar. Within the selected schools (on average 140-185 primary schools per country), fixed size clusters of 20 grade 6 pupils were randomly drawn for inclusion in the study[3]. High response rates ensured an average of 18.1 SACMEQ II pupil records per school (Lee et al. 2005: 215). After the study had been conducted, it was discovered that South Africa had been somewhat under-sampled, given the extremely high levels of clustering observed in this country. Apart from this issue, the dataset is generally thought to be of good quality (Lee et al. 2005: 215).

The nested nature of the data provides some econometric challenges. To start with, it is typically assumed in data of this kind that observations within clusters are not independent: the achievement of pupils enrolled at the same school is collectively influenced by the quality of education provided at that institution. To deal with the impact of this feature of the data on the standard errors of coefficient estimates (which will consequently be too small), an adjusted variance-covariance matrix must be estimated. The resulting parameter estimates are robust to any type of correlation between observations within the same cluster as well as to heteroskedasticity, and require only that the number of observations is large in relation to the number of clusters (which in the SACMEQ case it is) and that observations across clusters are independent (an assumption one is generally prepared to make, in the case of pupils nested within schools). OLS parameter estimates are consistent and asymptotically normal, and survey regression methods can be applied. As with simple OLS, however, a fundamental assumption underlying survey OLS regression is that explanatory variables are determined exogenously. If this assumption is violated and the explanatory variables are in fact influenced by the clustered observations, then inference is not reliable (Wooldridge 2001).

Another method for dealing with nested data in the context of student performance is advanced especially by analysts in the field of education. Hierarchical linear modelling (HLM, or multilevel analysis, as it is sometimes called) is used for estimation where nesting occurs at two or three levels. Apart from the implications of nested data for the estimation of OLS standard errors, Raudenbusch and Bryk (2002) highlight two additional econometric issues specific to multilevel data that the technique of hierarchical linear modelling addresses.

Aggregation bias occurs when a variable has differing impacts depending on the level at which it operates. For instance, socioeconomic status may impact on a child's academic achievement directly through influencing the size of the household budget for education; at the school level, however, socioeconomic status reflects the availability of school resources and perhaps also the quality of education received by pupils enrolled there. Examining postulated relationships at school-level would thus hide the relationship between socioeconomic status and achievement at the individual level. Hierarchical linear models model these effects separately, since students form the units of analysis at level 1 while schools are the units of analysis at level 2.

The second issue – and perhaps the most important one in terms of the focus of this paper – concerns the heterogeneity of regressions in a multilevel context. Schools are likely to convert resources into school outputs with varying efficiency, as a result of internal variation in teaching methods or resource management, and external variation in school-level factors including the socioeconomic status of the communities schools serve. Rather than operating on the

[3] In the case of the Seychelles, all 24 schools were included, given the small population size.

assumption that school-level variables affect student achievement uniformly across the sample, hierarchical linear models allow for random coefficients unique to schools. The hypothesis that explanatory variables have a "fixed effect" on the dependent variable can then be tested for validity, rather than being taken as given. School-level random effects can be incorporated in modelling the intercept or slope variables (or both) specified in the level 1 regression.

In this paper, results from both types of analysis will be presented. Survey regressions are estimated separately for the total sample, South Africa and the remaining countries. Following this discussion, the results of Van der Berg (2005) and Gustafsson's (2006) hierarchical linear modelling exercises on the SACMEQ II data for South African students are reviewed. Firstly, however, some descriptive statistics and univariate and bivariate analysis are presented.

Preliminary analysis

Tables 1a, b and c show descriptive statistics for all of the variables included in our regression analysis, divided into those that describe students and those that describe respectively mathematics teachers and schools, and disaggregated into figures for South Africa and the remaining SACMEQ countries.

With respect to the variables describing students, our SES (socioeconomic status) index was constructed following Lee et al. (2005), but without including parent's education. It is thus a z-score representing a composite of the quality of a student's home building (measured in terms of the materials used to construct floors, walls, the roof and the student's lighting source) and the student's possessions (having access to water, access to electricity, possessing a newspaper, magazine, radio, television, video cassette recorder, fridge, car, motorcycle, bicycle and table). SES is included as both a linear and quadratic explanatory variable in our regression analysis. Given a high number of missing values for the education of individual parents (as reported by their children), our parents' education variable is constructed as the maximum of mother and father's education. In addition to this variable, there is a dummy variable indicating whether or not the student's mother holds a tertiary qualification. The item index spans a range of 0-6, being a composite of dummy variables indicating whether a pupil owns a range of stationery items. The meal index numbers 0 to 9, describing how frequently the student eats (weekly basis, 3 meals a day). A dummy variable indicates whether a student lives with his/her parents, and another indicates whether the student's home contains more than 10 books. With respect to education variables of interest, dummy variables indicate whether a student has repeated any grades, whether the student uses his/her own textbook in class, whether he/she receives extra tuition and whether he/she does homework once a week or most days of the week (as opposed to less frequently). The average number of days a pupil is absent from school per month is also included. Finally, there are standard control variables, including gender and rural dummies and the student's age.

Regarding the variables describing schools, the mean level of parents' education is included alongside more conventional measures. The school facility index runs from 0 to 23, and is a composite of dummy variables indicating whether the school has electricity, access to water on tap or through a spring, bathroom facilities, a library, a school or community hall, a staffroom, an office for the principal, a playground or sports field, a fence or hedge around the school, a garden, a cafeteria/kiosk, and various types of school equipment (telephones, computers, projectors, etc). The school homework index runs from 0 to 3, averaging the frequency of homework given (1 = once per month, 2 = once per week, 3 = most days of the week). The proportion of students in grade 6 who have repeated grades is also included. Traditional school quality measures such as the pupil-teacher ratio, teacher's years' of experience, and dummies indicating whether or not the teacher has tertiary education or teacher training, are added.

Finally, there is a dummy indicating whether or not the school principal reports teacher absenteeism to be a problem.

Given the dominance of resources and parent's education in the literature presenting economic studies of student achievement, the starting point for this paper was an investigation of whether resource levels (in households and schools) and parents' education differed systematically between South Africa and the rest of the sample. For this task, t-tests were run on some key variables of interest. At the 1 percent level, South Africa proves to have significantly better school facilities, higher average teacher education, a greater proportion of teachers with tertiary education, teachers with more work experience, and a lower pupil-teacher ratio. South African pupils are also more likely to attend schools in urban areas – institutions one might expect to enjoy preferential access to good teachers and other school resources within a country. It is also usually the case that urban surroundings bring learning benefits through exposure to modern lifestyles, the written media, and English, the language of the test in most of the countries[4]. In addition, South African pupils enjoy more benefits at home than their peers do: their SES is higher, both parents are better educated, and more of their mothers have tertiary qualifications. However, despite this apparent abundance of resources, South African households do not seem to prioritise the education of children very highly. Students have less stationery with which to do their schoolwork and have fewer meals than their SACMEQ counterparts. Indeed, less than three quarters of South African children have breakfast before arriving at school (Wittenberg 2005: 11). While nutrition is not directly related to the educational process, it has been shown to have a significant impact on children's schooling outputs (McCoy, Barron & Wigton 1997, in Wittenberg 2005: 11).

To further investigate the relationships between test scores and both SES and parents' education, additional analysis was conducted. Figure 1 shows kernel densities of South African mathematics test scores versus the rest of the SACMEQ sample. Note that the modal score is lower than the one for the remaining countries, suggesting that South African performance on test scores is weak in aggregate. Figures 2 and 3 show lowess (locally weighted) regressions for mathematics score versus the SES score of pupils for South Africa and the other thirteen African countries. Each dot represents one pupil. Similar data, but this time showing the mean scores versus the mean SES of schools (each dot now representing a school), are also shown in Figures 4 and 5. Note how flat the South African lowess regressions are for lower levels of SES, in contrast to the other countries. This pattern can be observed even more clearly when the South African and non-South African lines are shown on the same graph, as reflected in Figures 6 (for individuals) and 7 (for schools). It is very apparent that at low levels of SES, mathematics test scores of both individuals and schools change very little as SES rises in South Africa, whilst there is a definite slope in the case of non-South African countries sampled. Consequently, the gap between the other countries and South Africa initially widens over a large range of SES values. At high levels of SES, the scores of South African pupils start to improve sharply and indeed overtake the other countries. This points to the extremely high levels of segmentation and inequality in the South African education system – largely a legacy of apartheid. The discussion on results of estimation in the next section picks this issue up again.

The lowess regressions do not show the weight of the distribution of individuals across SES scores. Figure 8 better illustrates this for individuals in each of the 14 SACMEQ countries, where the mean SES and mathematics score of each quintile of the population (arranged by SES) for each country are plotted. The values from which the graph was derived are shown in Tables 2 and 3. As can be seen, the five South African quintiles span a large range of SES values,

---

[4] Indeed, Lee et al. (2005) find a robust pattern of higher achievement in urban schools in the SACMEQ countries.

reflecting the underlying economic inequality. As is evident from this graph, the poorer South African quintiles perform rather poorly compared to the comparator countries. Starting from a low base, the slope measuring the relationship between SES and mathematics test scores is quite flat for the first three quintiles. In fact, in nine of the other African countries, even the bottom quintile performs better than the third South African quintile. It is only the steep slope of the SES-performance gradient between the higher quintiles that explains why more affluent South Africans perform relatively well in SACMEQ context. Yet note that even the top quintile of South Africans, arranged by SES score, does not perform as well as the top two quintiles in Seychelles, or even as well as the top quintile of Tanzanians – a group that is much poorer.

South African students also perform worse on the test than one would expect given the educational attainment of their parents. Tables 4 and 5 show the mean score by mother and father's education respectively. Note that for both mothers and fathers who have low levels of education, the mean mathematics score in South Africa is far below the southern African mean, whilst the differentials narrows at higher education levels of both the parents – even disappearing at higher level of father's education.[5]

Results of estimation

Before the discussion on model results begins, a few words on potential problems caused by endogeneity are in order. We would expect a student's ability to be a strong determinant of his/her mathematics test scores, although there is no explicit measure of this variable in the SACMEQII dataset. Econometric theory tells that such an omission will cause coefficient estimates on the explanatory variables that are positively correlated with ability to be upwardly biased. A similar problem exists for parents' willingness to help children with homework, a variable that is seldom ever controlled for in production function studies (Glewwe 2002). Typically researchers control only for parents' educational attainment, assuming that willingness to help with homework is positively related to parents' education, and indeed part of the impact thereof on child's education.

A potentially even more serious consequence of the omission, however, is that the ability of a student and his/her enrolment at school may be correlated in some poorly resourced African countries. While no longer a major problem in South Africa, access to schooling opportunities at even primary school level is scarce in some countries in the sample, and parents with more able children might make special arrangements to enable them to attend school (e.g. move to urban areas where there are more schools, or do additional household chores to enable their children to attend school). This implies that the children tested in SACMEQ II might be more able than average, and thus not representative of the population of children in their countries. Note, however, that ability is the issue here: If it was simply a matter of more educated parents or parents with higher socio-economic status having children that performed better, the coefficients in regression analysis would fully reflect this, as these variables are measured and included in most models.

Another factor contributing to the likelihood of the sample being less than completely representative is drop-out. While completion of primary schooling is universal throughout most of the rest of the world, impoverished African students who repeat grades are more likely to drop out early. If these students are less able, then this may also contribute to less reliable coefficient estimates. But once again, if drop-out was purely linked to SES, then the coefficients would still be unbiased as long as ability was not correlated with SES, which in African circumstances

[5] Note, however, that many father's were not present in the households in which many South African children found themselves, or that their education levels were unknown to the children. The table reflects only data for those parents for whom education levels were known.

appears to be a fairly acceptable assumption. The net result is that comparison between the model for South Africa (where access is less of a problem) and the rest of the sample may be slightly problematic. However, given the limitations of the SACMEQII dataset (and ability is seldom separately tested in most datasets), these are issues that we must live with. Further, we are encouraged by the results of Glewwe and Jacoby's education production function study for children in grades 7-10 in Ghana (discussed in Glewwe 2002). This work found that controlling for selection into schools made little difference to the regression results.

Survey regression

During the survey regression modelling process, we started with a mathematics score model including a large number of potentially important explanatory variables, and reduced these variables using a general-to-specific modelling approach. Once a general model emerged, we included South Africa interaction terms for a range of student and school variables informed by the results of our preliminary analysis. For this run, South Africa was chosen as the reference country. Subsequently, the same model without the interaction terms was run separately for South African students and the rest of the sample. In the latter case, Namibia was the reference country. Regression results are presented in Table 6.

The general model with SA interaction terms is contained in column 1. As might be expected, SES and parents' education are clearly important predictors of mathematics achievement. In addition the impact of the highest educational attainment of either parent, there is an additional benefit to children whose mothers have tertiary qualifications. Furthermore, there is school average parents' education also influences individual student achievement: students attending schools chosen by better educated parents, score better on the mathematics tests. This may be due to well-educated parents actively participating in school management issues, thus helping to ensure that their children receive good quality schooling. The impact of SES on test scores appears to be quadratic rather than linear, as suggested by preliminary analysis. HLM provides a vehicle for testing which factors drive this relationship, and the next section of this paper accordingly sheds some light on the apparent cause of the observed non-linearity. Note that there is an additional boost to test scores of children whose mothers have a degree; this suggests that the impact of mother's education on child's schooling performance is non-linear, since in the starting model there was no similar significant benefit for children with mothers who have completed secondary school. A very interesting finding is that South African children of any given SES perform significantly worse on the mathematics test than their SACMEQ peers do. This result formalizes the pattern observed earlier in the section reporting results of preliminary analysis on the bivariate relationship between SES and mathematics performance; the regressions indicate that this relationship is robust to the inclusion of control variables.

Turning to the remaining pupil-level variables, it is unsurprising to observe that private resources play a significant role in determining test scores. The item and meal indices, having more than 10 books at home and receiving additional tuition are all positively and significantly related to students' mathematics test scores. However, the extent to which a student benefits from such household resources is likely to be strongly influenced by whether he/she lives with his/her parents, which turns out to be a relatively large as well as highly significant positive determinant of test scores. Recent research on orphans in Africa by Case and Paxson (2002) suggests that the relationship of a child's care-giver to him/her matters a great deal for schooling. There are test score penalties for being older or being female. Children who have repeated grades also fare systematically worse, perhaps because this signals lower ability or learning problems. From a policy perspective, an important finding is that lack of own textbooks influences mathematics scores negatively.

Attention turns next to school-level factors. There appears to be a modest but clear benefit to attending a school with good facilities, with students from South Africa benefiting more for any given set of facilities. Rather than suggesting that the facilities themselves are driving better performance in South Africa, we would argue that schools with better facilities are most likely to be previously advantaged formerly white and Indian schools. Given the previous regime's policy of providing highly varying amounts of attention and funding to schools in different (racially based) education systems (see Case & Deaton 1999), these schools would have benefited from superior public resource bases and well-trained teachers, in addition to good management and a relatively affluent and well-educated parent body. Even now, after public resource levels across schools have largely been equalized, these schools continue to enjoy many of the benefits they did before, through being able to charge parents substantial school fees with which to augment publicly provided school resources.

However, analysis shows that this advantage in terms of performance cannot be fully explained by school or even private resources. Apart from the observed significance of physical resources, it appears that teacher resources also play an important role in influencing mathematics achievement. Receiving lessons from a teacher with tertiary education strongly boosts a child's test score, although this effect is much smaller in South Africa than in the other African countries. This lesser importance of teacher qualifications in South Africa can be explained by factors including great variability in the quality of tertiary education provided to South African teachers and weak management of teacher resources within the poorer schools. Other teacher variables that positively affect student test performance are being in possession of teacher training and having more years of teaching experience. There is also support for the hypothesis that the pupil-teacher ratio matters, with students in smaller classes performing better on the mathematics tests. But this effect is not very large – the regression indicates that the pupil-teacher ratio would have to change very much to greatly affect school outcomes.

Apart from the classroom factors that are traditionally highlighted in the education production function literature, there are a few interesting ones that emerge as significant determinants of test performance in the SACMEQ countries. The average frequency of homework given by a school, the proportion of grade 6 students at the school who have repeated one or more grades, and a principal-reported problem of teacher absenteeism may be viewed as variables strongly correlated with school quality. Good quality schools are likely to be more concerned with student progress – thus giving homework more frequently and trying to ensure that all students meet the standard required to pass grades – and with good management of school resources – thus being less likely to tolerate unwarranted teacher absenteeism. In our regression all three proxy variables are highly significant determinants of test scores, and have the expected signs. Interestingly, the test score premium for attending a school that gives students homework frequently is much larger for South African students, while their penalty for attending a school that has a larger proportion of students who have repeated grades is much larger. Once again, these findings are more likely to be a reflection of the cleavage between the formerly Model-C (largely white and Indian) schools on the one hand and schools previously falling under the authority of the Department of Education and Training (i.e. the poorly resourced and managed black schooling system) on the other.

Finally, it bears mentioning that after controlling for the entire set of explanatory variables included here, students in each of the other 13 SACMEQ countries perform better on their mathematics tests than South African students do. All country dummies are positive and also significant (South Africa is the reference country). Dummies range from 36 to 177, very large values when compared to the standard deviation of scores of 100. This is a surprising and disturbing finding for a country that is amongst the most affluent in the sample – a country that is home to a sophisticated economy, but in which a skills shortage is currently viewed as an obstacle to achieving more rapid growth.

Another way to compare South African performance compared to the African norm is to estimate expected values for the South African scores based on the full model, excluding country dummies, reflected in column 4 of Table 6. The predicted values for South Africa from this regression is compared for different quintiles of the distribution in Table 7. As is apparent, the South African conditional mean test score is then 516, compared to an actual mean of 486. The top South African quintile's performance is slightly above expectations, whilst the performance of the other quintiles is consistently much worse, about a third of a standard deviation below expectations.

We turn now to regression models estimated separately for South Africa and the remaining SACMEQ countries. Given the large overlap between the total sample and the sample excluding South Africa, there are no surprises in the estimation results for the model in column 3 of Table 6. Note that Namibia is the reference country for this model. Turning to the South Africa model in column 2 of the same table, it is interesting to note that some variables that enter as significant determinants of test scores in the rest of the SACMEQ sample do not seem to matter for South African mathematics test scores. Neither parents' educational attainment nor the frequency with which students do homework significantly influence performance on the mathematics test[6] in South Africa. Furthermore, none of the three teacher quality measures have a significantly positive impact on test scores. In the case of the teacher education variables, this is likely to be due to the highly variable quality of training received by the teachers, who would have obtained their education during the apartheid years. There are also some variables that appear to have substantially different impacts on test scores in South African than in the rest of the SACMEQ sample. The impact of the item index on test scores is larger, as is having more than 10 books at home, or owning a textbook. In the South African case we observe a positive effect of attending a school with well-educated parents that is bigger than it is in the rest of the region. However, none of these differences appear to be statistically significant; only the variables captured as South African interaction terms in the previous paragraphs have significantly different impacts than they do in the remainder of the sample. Finally, note that in South Africa it seems that there are no test score penalties for being in a school with a higher pupil-teacher ratio, living in a rural area, being female or being older.

Hierarchical linear modelling

There are a number of hierarchical linear model types that can be used to answer different questions about schooling outputs. However, here we will only consider only two types: one-way ANOVA and the means-and-slopes-as-outcomes model. The results of HLM estimation contained in Van der Berg (2005) and Gustafsson (2006) form the basis of discussion in this section.

In a one-way ANOVA model, there is only one fixed effect: the overall sample mean of the outcome variable in level 1, β_{0j} (i.e. the weighted average mathematics score, in our analysis). Note that subscript i refers to students (level 1 units) while j refers to schools (level 2 units). The level 1 equation models achievement in mathematics by including this estimated parameter and a random effect for each student (assume $r_{ij} \sim N(0,\sigma^2)$):

$$Y_{ij} = \beta_{0j} + r_{ij}$$

[6] Each of the four variables in these two sets of variables is tested for individual and joint significance.

In HLM, the level 2 equations model the coefficients taken as given in level 1, using variation in level 2 (school-level) factors. In the ANOVA set-up, β_{0j} is modelled as the sum of the grand mean mathematics achievement, γ_{00}, and a random effect for each school in the sample, u_{0j}. We assume that $u_{0j} \sim N(0, \tau_{00})$, where τ_{00} is the parameter variance.

$$\beta_{0j} = \gamma_{00} + u_{0j}$$

Expressing the level 1 regression equation as the sample mean of the outcomes for all n_j students in the total j schools, the total (mixed) model can be stated as:

$$\overline{Y}_{.j} = \gamma_{00} + u_{0j} + \overline{r}_{.j}$$

The usefulness of the ANOVA exercise derives from the total variance in mathematics scores being partitioned into a component attributable to factors driving performance within schools and another attributable to factors determining differentials in performance between schools. The latter is measured by the intra-class correlation coefficient, and represents the theoretically maximum variance in the dependent variable capable of being explained by level 2 factors. This is computed as:

$$\rho = \frac{\tau_{00}}{\tau_{00} + \sigma^2}$$

In South Africa, the SACMEQ II rho value for mathematics scores was estimated to be 0.64. This was much higher than comparable values for the rest of the sample, attesting to the highly unequal, formerly racially segmented schooling system operating in this country.

In the case of the means-and-slopes-as-outcomes model, both the intercept term and the slope coefficients in the level 1 equation are modelled through adding regressors to explain these at level 2. Supposing that one regressor is added to each equation and that random effects are allowed for, the model set-up would be:

$$Y_{ij} = \beta_{0j} + \beta_{1j} X_{ij} + r_{ij}$$
$$\beta_{0j} = \gamma_{00} + \gamma_{01} W_1 + u_{0j}$$
$$\beta_{1j} = \gamma_{10} + \gamma_{11} W_1 + u_{1j}$$

One final concept that needs to be introduced here is the reliability ratio. This reflects what the average reliability of random intercepts and slopes would have been had OLS been used, and is computed as follows:

$$\lambda(\hat{\beta}_q) = \frac{1}{J}\sum_{j=1}^{J} \tau_{qq} / (\tau_{qq} + v_{qqj}) \text{ for each q = 0, … Q.}$$

Here v_{qqj} represents the sampling variance of $\hat{\beta}_{qj}$. When constructing a hierarchical linear model, one generally includes random effects in each equation. A high reliability ratio then suggests that the random effects should be retained, while a low reliability ratio (<0.10) indicates that fixed effects are more appropriate than either random effects or non-randomly varying effects. Reliabilities depend on the extent of variation between groups in the true underlying parameters, and on the precision with which each group's regression equation is estimated. The reliability ratio is thus close to 1 when the group means vary substantially across schools, or in large samples (Raudenbush & Bryk 2002: 46).

Van der Berg (2005) models SACMEQ II mathematics test scores using predominantly individual-level factors. However, the intercept term and the slope coefficient on individual SES in level 1 were modelled at level 2 by including mean school SES and random effects as regressors (i.e. treating these effects as varying non-randomly). Where values were missing for any variable in the case of all observations from a given school, all observations for the school were dropped from analysis, reducing the sample. The variables were entered raw. Results of estimation are presented in Table 8.

Modelling the impact of SES within an HLM framework sheds much light on the observed non-linear nature of the relationship between test scores and SES. Apart from having a positive level effect on test scores (through the intercept), mean school SES influences achievement indirectly through interaction with individual SES. The observed non-linearity between individual SES and test scores can be explained in terms of community-level factors compounding the impact of a student's family background. In seeking to understand the implications of this finding, it is illuminating to note that the correlation between individual SES and mean school SES in South Africa is fourth highest in the sample of SACMEQ countries, and is particularly high relative to the average values for the rest of the SACMEQ sample when calculated separately for each of the poorest three quintiles (measured on the basis of a constructed SES measure). This reflects the fact that South Africa remains a highly stratified society, with socioeconomic class currently increasingly replacing race as the dividing line. Our analysis suggests that to the extent that schools in better resourced neighbourhoods provide superior quality education (as one would expect they might), poor South African students are less likely to escape their home background disadvantage than are the bulk of their peers in the southern African region. This is particularly troubling giving the apparent flatness of the relationship between individual SES and mathematics scores for the poorest three quintiles (as noted earlier), which suggests that 60 percent of students are trapped in situations where the obstacles to academic achievement are so great that even a relatively large range of SES makes no apparent difference to their test scores.

Analysing the model output regarding random school effects, Van der Berg (2005) notes that standard deviations are large. In particular, the model confirms that there is a large degree of variation in the relationship between SES and mathematics scores at the school level. We digress briefly, to turn to additional survey regression results obtained in Van der Berg (2005). The highly varying relationship between SES and test scores in South Africa was examined more closely by dropping first the most affluent decile of schools, and then the next most affluent decile. The purpose of this exercise was to isolate and remove formerly white and Indian schools, to test whether the data generating process in these schools was the same as that operating in the historically black and coloured schools. In the case of the mathematics test score model, the SES coefficient declined dramatically in size and significance during the sample reductions; in the fully reduced (i.e. 80 percent) sample, SES does not appear to play a significant role in shaping mathematics achievement. Indeed, it is much more difficult to identify which variables do drive student performance in the poorer schools, as the low coefficients of determination for the reduced sample models in Van der Berg's (2005) analysis show. This is the story that Crouch and Mabogoane (1998) tell, in their aptly titled article "When the residuals matter more than the coefficients".

The strong explanatory power of the HLM model is reflected in a decomposition of the remaining variance in the model relative to the model with no regressors (i.e. one-way ANOVA).

There is a decline in unexplained individual test score variance between schools of 70 percent, and a drop in variance within schools of 6 percent. The fact that variance reduction occurred mainly through adding variables that differ across schools is line with what we know about persistence in school-level private resource bases as well as school operation and outcomes following the end of apartheid.

Gustafsson (2006) models assessment scores using HLM for South Africa from the SACMEQ data, but distinguishes advantaged schools – which he defines as schools in the top quintile of the SES distribution – from disadvantaged schools. Like Van der Berg (2005), he assumes that this largely eliminates from the disadvantaged sample historically white and Indian schools, which generally perform much better in school assessment tests. He starts by modelling both mathematics and reading scores for the full sample and the separate advantaged and disadvantaged sub-samples, using only home background factors, viz. SES and parent education, to ascertain what part of the variance these explain and how much variance is left to be potentially explained by policy factors. In the full sample, he finds that SES and parent education explain about half of the between-school variance. However, because the segmentation by SES level itself explains much of the variance, these variables explain much less of the between-school variance in the sub-samples. He points out that it is particularly striking how little (only 7 percent) of the between-school variance in mathematics performance in the historically disadvantaged sub-sample is explained by these household characteristics.

His full model is run only for reading scores and only for the sub-sample of disadvantaged schools. It allows for some school-level (level 2) effects in the intercept term, and he models this as a random effect, i.e. allowing the effect of individual schools to vary. He also models two level 1 slopes as being affected by school level variables, viz. individual SES plays a role in interaction with class size, and school's mean SES (i.e. school affluence) affects the impact of gender on reading outcomes. Both these effects are modelled as fixed rather than random, based on reliability ratios.

One conclusion from this analysis is that large class size affects how SES is translated into test scores, and he then postulates that schools with very high class sizes do not translate high SES into improved output scores.[7] Further analysis of raw data show this to be the case, but the aggregate effect on test scores is relatively small. A second major conclusion that Gustafsson draws is that inputs at school level play only a small role in explaining between-school variance, a conclusion that is strengthened by the comparison he draws with the Brazilian case. This supports earlier analyses, including that of Crouch and Mabogoane (1998), Van der Berg and Burger (2002) and Van der Berg (2005). As in these earlier studies, he ascribes this to how inefficiently these school inputs are utilised. Gustafsson (2005 and 20006) tries to relate his conclusions regarding school performance to the cost of interventions. He concludes that cost-effective interventions may include in-service training of teachers, access to text books, and problems reported by principals regarding teachers arriving at school late (which he points out may be part of a bigger motivation problem) (Gustafsson 2005: 28-32).

Conclusion

This paper is concerned with trying to understand why South African mathematics performance is explained to such a large degree between schools, and why it is so poor in regional context. In regional context, South Africa performs surprisingly poorly. At lower SES levels, South Africa is well behind the average for the SACMEQ countries. Nine of the countries in the sample start

[7] Note that class size refers to the actual number of children in a given class. It is not necessarily the same as the pupil-teacher ratio, as it depends not only on teacher availability or the pupil-teacher ratio for the whole school, but also on how teachers are distributed and applied within a school.

from a higher base in terms of test scores, and performance climbs in a more linear way with SES. Even the steep improvement in South African test scores observed for the uppermost quintiles is not particularly impressive in regional context; students in Tanzania and the Seychelles achieve better results. This pattern, of relative SES advantage not translating into good test scores, is echoed with respect to South African parent education levels. Indeed, having better school resources (including higher teacher education) does not appear to make the difference one might expect it would in South Africa.

We have shown that there is an extremely high degree of segmentation in the South African schooling system, with students in the two most affluent quintiles exhibiting the positive relationship between test performance and socio-economic status that education production function theory predicts. It is likely that this results from the much better educational performance in historically white and Indian schools, which is also apparent in other studies. Although these schools are now open to all race groups, their student bodies are still dominated by the more affluent, regardless of race. By contrast, there is no visible relationship between SES for students in the poorest three quintiles – a group that performs very poorly in regional context. The slope is almost completely flat over this wide range of SES, pointing to problems of widespread inadequacy in resource management in the dysfunctional (formerly black) part of South Africa's schooling system. There is a strong need for policy attention at this end of the spectrum, given that the least affluent 60 percent of South African students are not particularly poor in regional context. Given the observed pervasive and enormous inefficiency in dysfunctional South African schools, throwing more money at the school quality problem is very unlikely to solve the problem. However, the relative "over-performance" observed by Crouch and Mabogoane (1998) in some poor schools should provide policymakers with encouragement as well as an avenue for learning policy lessons that can be applied throughout the rest of the system.

The regression analyses performed here, and indeed also earlier work referred to, support a view that South African mathematical performance in this SACMEQ study is worse than expected relative to other African countries because of the poor ability of a large part of the school system to transform advantages in school resources (including pupil-teacher ratio and teacher education levels), parent education and socio-economic status into systematically better school mathematics performance.

There is a widely held view in South Africa that this problem largely relates to the systematic disadvantages that most black and many coloured schools experienced under apartheid. Indeed, this must be at least an important, if not the dominant, historical explanation for the extremely weak results of the SA education system. The evidence presented shows that this disadvantage still applied in 2001, and other evidence shows that the school system's performance has not improved since[8] – there is still a systematic underperformance in large parts of the school system. This would conform with systematic underperformance of historically black schools, and indeed other studies also conform with such an explanation. However, as Crouch and Mabogoane (1998) indicate, it is possible to identify poor schools who perform above their conditional expectations, and understanding performance in such over-performing schools is crucial. The key to their efficiency, these authors argue, lies in school management and culture of learning processes, and in "better information and incentive and control systems that.for the optimal deployment of ... resources" (Crouch and Mabogoane 1998: 2, 4).

[8] Matriculation results, particularly in terms of higher quality of matric passes, indicate that there has been little improvement in the numbers achieving passes, exemptions or A-aggregates, and these numbers are not keeping up with the growth of the matriculation-aged cohort. Assessment tests performed annually at different grades in the Western Cape show weak performance to remain endemic in a large part of this province's schools.

Production function analyses of South African schools have tended to arrive at the conclusion that the continuing poor performance of this dominant part of the South African school system may result from a deeper malaise, which resource shifts have not yet been able to overcome: endemic inefficiency of a large part of the school system. However, in such studies it is more difficult to show the inefficiency in the poorer South African schools, because such inefficiency is strongly correlated with SES. However, the comparison with other African countries allows us to draw stronger conclusions about efficiency in poor South African schools. Also, by focusing on the relationship between SES and performance, this paper presents evidence that SES only appears to have an impact in South Africa at a high threshold level.

The most important implication for policy is that management and efficiency issues in schools deserve greater attention, and that greater resource inputs by themselves are unlikely to greatly improve performance of weaker performing schools. Also, the availability of the SACMEQ data shows that the problem of weak performance in matric, which has been the subject of most of the earlier production function analysis in South Africa, originates much earlier. It is precisely the lack of such data that potentially blinds policy makers to the deep-seated malaise of the school system, and to the fact that directing interventions at the secondary school level may be too late, a point that is reflected by weak and varying matriculation results (Renault-Smith 2006). Systematic assessment at primary school level and early in the secondary school cycle would reveal deficiencies much more clearly and allow policy makers to better assess possible intervention strategies than has hitherto been the case. Moreover, such information itself can potentially be used as source of information to empower communities as well as schools, thus providing a resource for getting to grips with the systematic underperformance of a large part of the school system that continues to deprive most poor South Africans from taking their rightful place in the labour market.

REFERENCES

ANGRIST, J.D. & LAVY, V. 1999. Using Maimonides' rule to estimate the effect of class size on scholastic attainment. *Quarterly Journal of Economics* 114(2): 533-75.

CASE, A. & DEATON, A. 1999. School inputs and educational outcomes in South Africa. *Quarterly Journal of Economics* 114(3): 1047-84.

CASE, A., PAXSON, C. & ABLEIDINGER, J. 2002. *Orphans in Africa*. NBER Working Paper No. 9213. [Online.] Available: www.nber.org

CROUCH, LUIS & MABOGOANE, THABO. 1998. When the residuals matter more than the coefficients: An educational perspective. *Studies in Economics and Econometrics* 22(2): 1-14

GUSTAFSSON, MARTIN. 2005. *The relationships between schooling inputs and outputs in South Africa: Methodologies and policy recommendations based on the 2000 SACMEQ dataset.* Paper presented to the SACMEQ conference, Paris, September.

GUSTAFSSON, MARTIN A. 2006. *School production modelling to strengthen government monitoring programmes in developing countries.* Master's thesis in Economics. Stellenbosch: University of Stellenbosch

HUMAN SCIENCES RESEARCH COUNCIL (HSRC). 2005. *TIMSS 2003 results*. [Online.] Available: http://www.hsrc.ac.za/research/programmes/ESSD/timss2003/index.html

FULLER, B. 1985. *Raising school quality in developing countries: what investments boost learning.* World Bank Discussion Paper 2. World Bank: Washington, D.C.

HANUSHEK, E. 1986. The economics of schooling: production and efficiency in public schools. *Journal of Economic Literature* 24(3): 1141-1177.

HANUSHEK, E. 1995. Interpreting recent research on schooling in developing countries. *World Bank Research Observer* 10(2): 227-46. August.

KRUEGER, A.B. 1999. Experimental estimates of education production functions. *Quarterly Journal of Economics* 114(2): 497-532.

LEE, V.E., ZUZE, T.L., & ROSS, K.N. 2005. School effectiveness in 14 Sub-Saharan countries: Links with 6th graders' reading achievement. *Studies in Educational Evaluation* 31: 207-246.

MOLOI, MESHACK Q. 2005. *Mathematics achievement in South Africa: A comparison of the official curriculum with pupil performance in the SACMEQ II Project.* Paper presented to the SACMEQ conference, Paris, September.

RAUDENBUSH, S.W. & BRYK, A.S. 2002. *Hierarchical linear models: Applications and data analysis methods*. Second edition. California: Sage Publications.

RENAULT-SMITH, JENNIFER. 2006. Matric improvement programmes. Ch.16 in: Reddy, Vijay (Ed.). 2006. *Marking Matric: Colloquium proceedings.* Pretoria: Human Sciences Research Council: 228-238

TAYLOR, N., MULLER, J. & VINJEVOLD, P. 2003. *Getting schools working*. Cape Town: Pearson Education.

VAN DER BERG, SERVAAS. 2005. *How effective are poor schools? Poverty and educational outcomes in South Africa*. Paper presented to the SACMEQ conference, Paris, September.

VAN DER BERG, SERVAAS; & BURGER, RONELLE. 2002. Education and socio-economic differentials: A study of school performance in the Western Cape. *South African Journal of Economics* 71(3): 496 –522.

WITTENBERG, MARTIN. 2005. *The school day in South Africa*. Paper to the Economics Society of South Africa conference, Durban, September.

WOOLDRIDGE, J.M. 2001. First edition. *Econometric analysis of cross section and panel data*. Cambridge, Mass: MIT Press.

Table 1a: Summary statistics – Pupil-level variables

Country	Statistics	Age	Male	Stay with parents	More than 10 books	Meal index	Pupil has repeated grades
BOTSWANA	Mean	12.85	0.49	0.82	0.28	7.72	0.31
	Std Dev.	1.15	0.50	0.38	0.45	1.86	0.46
	Sample	3322	3322	3322	3322	3322	3322
KENYA	Mean	13.79	0.50	0.81	0.30	8.16	0.64
	Std Dev.	2	0.50	0.39	0.46	1.57	0.48
	Sample	3299	3299	3299	3299	3299	3299
LESOTHO	Mean	14	0.44	0.78	0.20	7.74	0.61
	Std Dev.	1.83	0.50	0.42	0.40	1.94	0.49
	Sample	3155	3155	3155	3155	3155	3155
MALAWI	Mean	12.21	0.52	0.85	0.11	8.34	0.66
	Std Dev.	2.15	0.50	0.36	0.32	1.16	0.47
	Sample	2333	2333	2333	2333	2333	2333
MAURITIUS	Mean	10.20	0.52	0.94	0.58	8.62	0.19
	Std Dev.	0	0.50	0.24	0.49	1.05	0.39
	Sample	2945	2945	2945	2945	2945	2945
MOZAMBIQUE	Mean	15	0.60	0.80	0.23	7.60	0.78
	Std Dev.	1.90	0.49	0.40	0.42	1.91	0.41
	Sample	3177	3177	3177	3177	3177	3177
NAMIBIA	Mean	13.64	0.48	0.73	0.26	7.73	0.54
	Std Dev.	1.87	0.50	0.44	0.44	1.88	0.50
	Sample	5048	5048	5048	5048	5048	5048
SEYCHELLES	Mean	11.51	0.50	0.87	0.52	7.27	0.10
	Std Dev.	1	0.50	0.34	0.50	1.99	0.30
	Sample	1484	1484	1484	1484	1484	1484
SWAZILAND	Mean	13.58	0.48	0.85	0.28	7.96	0.59
	Std Dev.	1.65	0.50	0.36	0.45	1.58	0.49
	Sample	3139	3139	3139	3139	3139	3139
TANZANIA	Mean	14.66	0.48	0.89	0.22	7.58	0.23
(excl. Zanzibar)	Std Dev.	1	0.50	0.32	0.42	1.75	0.42
	Sample	2854	2854	2854	2854	2854	2854
UGANDA	Mean	14	0.55	0.73	0.28	6.95	0.53
	Std Dev.	1.82	0.50	0.45	0.45	2.29	0.50
	Sample	2642	2642	2642	2642	2642	2642
ZAMBIA	Mean	13.67	0.52	0.83	0.23	7.70	0.52
	Std Dev.	1.68	0.50	0.38	0.42	1.84	0.50
	Sample	2611	2607	2611	2611	2611	2611
ZANZIBAR	Mean	14.57	0.48	0.94	0.09	8.20	0.28
	Std Dev.	1.56	0.50	0.23	0.29	1.75	0.45
	Sample	2514	2514	2514	2514	2514	2514
SOUTH AFRICA	Mean	13	0.47	0.78	0.35	7.53	0.42
	Std Dev.	1.60	0.50	0.41	0.48	1.97	0.49
	Sample	3163	3163	3163	3163	2696	3163
Total	Mean	13.36	0.50	0.82	0.28	7.81	0.48
	Std Dev.	1.99	0.50	0.38	0.45	1.82	0.50
	Sample	41686	41682	41686	41686	41219	41686
non-SA	Mean	13.41	0.50	0.82	0.27	7.83	0.48
	Std Dev.	2.01	0.50	0.38	0.44	1.81	0.50
	Sample	38523	38519	38523	38523	38523	38523

Country	Statistics	Homework given once per week	Homework given most days of the week	Pupil has extra tuition	Own textbook	Days absent per month	Item index
BOTSWANA	Mean	0.39	0.42	0.28	0.80	0.41	5.13
	Std Dev.	0.49	0.49	0.45	0.40	1.08	1.06
	Sample	3322	3322	3322	3322	3322	3322
KENYA	Mean	0.25	0.69	0.60	0.23	1.96	4.93
	Std Dev.	0.44	0.46	0.49	0.42	2.93	1.29
	Sample	3299	3299	3299	3299	3299	3299
LESOTHO	Mean	0.38	0.49	0.31	0.46	1.32	5.16
	Std Dev.	0.48	0.50	0.46	0.50	2.09	1.10
	Sample	3155	3155	3155	3155	3155	3155
MALAWI	Mean	0.57	0.20	0.09	0.57	2.00	4.42
	Std Dev.	0.50	0.40	0.28	0.50	2.79	1.09
	Sample	2333	2333	2333	2333	2333	2333
MAURITIUS	Mean	0.14	0.81	0.78	0.96	1.82	5.46
	Std Dev.	0.34	0.39	0.42	0.20	2.48	1.02
	Sample	2945	2945	2945	2945	2945	2945
MOZAMBIQUE	Mean	0.38	0.59	0.41	0.58	2.75	4.26
	Std Dev.	0.49	0.49	0.49	0.49	2.70	1.35
	Sample	3177	3177	3177	3177	3177	3177
NAMIBIA	Mean	0.24	0.67	0.22	0.48	1.49	4.76
	Std Dev.	0.43	0.47	0.41	0.50	2.35	1.31
	Sample	5048	5048	5048	5048	5048	5048
SEYCHELLES	Mean	0.44	0.51	0.21	0.76	0.86	5.16
	Std Dev.	0.50	0.50	0.41	0.43	1.46	1.35
	Sample	1484	1484	1484	1484	1484	1484
SWAZILAND	Mean	0.47	0.43	0.13	0.75	0.83	5.31
	Std Dev.	0.50	0.50	0.34	0.43	1.57	0.88
	Sample	3139	3139	3139	3139	3139	3139
TANZANIA	Mean	0.22	0.54	0.58	0.07	2.15	4.63
(excl. Zanzibar)	Std Dev.	0.42	0.50	0.49	0.25	3.55	1.32
	Sample	2854	2854	2854	2854	2854	2854
UGANDA	Mean	0.36	0.45	0.46	0.12	1.90	4.37
	Std Dev.	0.48	0.50	0.50	0.33	1.99	1.71
	Sample	2642	2642	2642	2642	2642	2642
ZAMBIA	Mean	0.48	0.27	0.28	0.13	2.52	3.81
	Std Dev.	0.50	0.45	0.45	0.33	3.10	1.66
	Sample	2611	2611	2611	2611	2611	2611
ZANZIBAR	Mean	0.46	0.19	0.31	0.05	1.98	4.62
	Std Dev.	0.50	0.39	0.46	0.22	3.11	1.47
	Sample	2514	2514	2514	2514	2514	2514
SOUTH AFRICA	Mean	0.32	0.54	0.30	0.41	1.64	4.27
	Std Dev.	0.47	0.50	0.46	0.49	2.83	1.99
	Sample	3163	3163	3163	3163	3163	3163
Total	Mean	0.35	0.51	0.36	0.46	1.68	4.74
	Std Dev.	0.48	0.50	0.48	0.50	2.60	1.43
	Sample	41686	41686	41686	41686	41686	41686
non-SA	Mean	0.35	0.50	0.36	0.46	1.68	4.78

	Std Dev.	0.48	0.50	0.48	0.50	2.58	1.36
	Sample	38523	38523	38523	38523	38523	38523

Country	**Statistics**	**Mother has a degree**	**Parents' education**	**SES**	**Maths score**
BOTSWANA	Mean	0.07	3.80	0.10	512.87
	Std Dev.	0.25	1.83	0.90	82.14
	Sample	3322	3058	3322	3321
KENYA	Mean	0.06	4.39	-0.38	563.25
	Std Dev.	0.23	2	0.67	87.47
	Sample	3299	3171	3299	3296
LESOTHO	Mean	0.05	4	-0.31	447.18
	Std Dev.	0.22	1.58	0.63	60.36
	Sample	3155	2811	3155	3144
MALAWI	Mean	0.02	3.50	-0.47	432.93
	Std Dev.	0.14	1.57	0.69	56.12
	Sample	2333	2219	2333	2323
MAURITIUS	Mean	0.02	4.20	1.58	584.58
	Std Dev.	0.14	1	0.47	139.89
	Sample	2945	2765	2945	2870
MOZAMBIQUE	Mean	0.02	4	-0.31	530.01
	Std Dev.	0.13	1.64	0.84	56.73
	Sample	3177	2822	3177	3136
NAMIBIA	Mean	0.06	4.19	-0.13	430.86
	Std Dev.	0.24	1.81	0.97	83.60
	Sample	5048	4113	5048	4990
SEYCHELLES	Mean	0.06	4.90	1.40	554.33
	Std Dev.	0.25	1	0.49	107.22
	Sample	1484	1215	1484	1482
SWAZILAND	Mean	0.08	4.34	0.13	516.54
	Std Dev.	0.28	1.84	0.83	67.35
	Sample	3139	2863	3139	3138
TANZANIA	Mean	0.05	3.61	-0.52	522.40
(excl. Zanzibar)	Std Dev.	0.21	2	0.69	86.36
	Sample	2854	2777	2854	2849
UGANDA	Mean	0.06	4	-0.63	506.28
	Std Dev.	0.24	1.79	0.64	108.43
	Sample	2642	2473	2642	2619
ZAMBIA	Mean	0.05	4.53	-0.33	435.21
	Std Dev.	0.21	1.57	0.84	72.86
	Sample	2611	2465	2611	2590
ZANZIBAR	Mean	0.03	3.59	-0.23	478.13
	Std Dev.	0.17	1.83	0.87	62.73
	Sample	2514	2409	2514	2459
SOUTH AFRICA	Mean	0.10	4	0.58	486.15
	Std Dev.	0.30	1.83	1.00	109.06
	Sample	3163	2602	3163	3135
Total	Mean	0.05	4.02	-0.01	496.80
	Std Dev.	0.22	1.73	1.00	99.65
	Sample	41686	37763	41686	41352
non-SA	Mean	0.05	3.99	-0.06	497.67
	Std Dev.	0.22	1.72	0.98	98.79

	Sample	38523	35161	38523	38217

Table 1b: Summary statistics – Teacher-level variables

Country	Statistics	Teacher has a degree	Teacher has teaching training	Teacher's years of experience
BOTSWANA	Mean	0.06	0.96	10.78
	Std Dev.	0.23	0.21	6.61
	Sample	395	395	395
KENYA	Mean	0.00	0.96	14.26
	Std Dev.	0.06	0.20	8.12
	Sample	268	268	263
LESOTHO	Mean	0.05	0.91	17.75
	Std Dev.	0.21	0.28	10.01
	Sample	231	231	231
MALAWI	Mean	0.00	0.88	7.80
	Std Dev.	0.00	0.33	5.47
	Sample	146	146	142
MAURITIUS	Mean	0.03	1.00	22.53
	Std Dev.	0.17	0.05	8.39
	Sample	417	417	417
MOZAMBIQUE	Mean	0.00	0.84	9.84
	Std Dev.	0.00	0.37	6.91
	Sample	328	328	323
NAMIBIA	Mean	0.10	0.97	11.95
	Std Dev.	0.30	0.17	8.14
	Sample	326	326	326
SEYCHELLES	Mean	0.05	0.98	11.47
	Std Dev.	0.22	0.14	10.31
	Sample	45	45	45
SWAZILAND	Mean	0.10	0.94	9.76
	Std Dev.	0.30	0.25	7.02
	Sample	171	171	169
TANZANIA	Mean	0.00	0.98	12.42
(excl. Zanzibar)	Std Dev.	0.00	0.14	7.35
	Sample	201	201	198
UGANDA	Mean	0.05	0.72	6.52
	Std Dev.	0.22	0.45	5.01
	Sample	168	168	127
ZAMBIA	Mean	0.01	0.97	11.56
	Std Dev.	0.08	0.16	9.04
	Sample	290	290	287
ZANZIBAR	Mean	0.00	0.92	11.04
	Std Dev.	0.00	0.28	8.86
	Sample	180	180	173
SOUTH AFRICA	Mean	0.26	0.95	13.64
	Std Dev.	0.44	0.23	7.54
	Sample	187	187	181
Total	Mean	0.05	0.94	13.18
	Std Dev.	0.22	0.24	9.09
	Sample	3353	3353	3277
non-SA	Mean	0.04	0.94	13.15
	Std Dev.	0.19	0.24	9.17
	Sample	3166	3166	3096

Table 1c: Summary statistics – School-level variables

Country	Statistics	Proportion of students who have repeated	Average frequency of homework given	Pupil-teacher ratio	School facility index	Average parents' education
BOTSWANA	Mean	0.31	2.16	28.33	10.75	3.79
	Std Dev.	0.14	0.40	4.49	3.21	1.00
	Sample	170	170	170	170	170
KENYA	Mean	0.64	2.61	33.36	8.23	4.40
	Std Dev.	0.20	0.34	9.24	3.11	0.91
	Sample	185	185	184	184	185
LESOTHO	Mean	0.61	2.34	53.85	6.64	3.68
	Std Dev.	0.19	0.44	18.54	2.30	0.76
	Sample	177	177	177	177	177
MALAWI	Mean	0.66	1.90	69.97	4.52	3.51
	Std Dev.	0.24	0.75	30.08	2.21	0.96
	Sample	140	140	140	140	140
MAURITIUS	Mean	0.19	2.74	24.50	15.38	4.19
	Std Dev.	0.10	0.36	13.72	2.08	0.60
	Sample	153	153	153	153	153
MOZAMBIQUE	Mean	0.78	2.56	51.32	7.37	3.52
	Std Dev.	0.14	0.22	36.25	4.24	0.69
	Sample	176	176	168	168	176
NAMIBIA	Mean	0.54	2.54	31.47	10.65	4.19
	Std Dev.	0.19	0.36	7.22	5.92	0.90
	Sample	270	270	270	270	269
SEYCHELLES	Mean	0.10	2.46	16.57	17.45	4.91
	Std Dev.	0.08	0.27	3.92	1.56	0.36
	Sample	24	24	24	24	24
SWAZILAND	Mean	0.59	2.30	35.11	9.13	4.36
	Std Dev.	0.17	0.59	6.69	3.58	0.88
	Sample	168	168	168	168	168
TANZANIA	Mean	0.23	2.15	47.06	5.89	3.61
(excl. Zanzibar)	Std Dev.	0.24	0.54	19.81	2.37	0.88
	Sample	181	181	181	181	181
UGANDA	Mean	0.53	2.19	57.98	7.12	3.91
	Std Dev.	0.21	0.48	24.74	3.80	0.89
	Sample	163	163	163	163	163
ZAMBIA	Mean	0.52	1.93	53.74	7.27	4.53
	Std Dev.	0.20	0.56	40.36	4.19	0.82
	Sample	173	173	169	169	173
ZANZIBAR	Mean	0.28	1.71	34.97	6.36	3.60
	Std Dev.	0.15	0.46	8.53	2.70	0.87
	Sample	145	145	145	145	145
SOUTH AFRICA	Mean	0.42	2.36	36.53	12.55	4.52
	Std Dev.	0.23	0.41	6.45	6.66	1.03
	Sample	169	169	167	167	169
Total	Mean	0.48	2.31	40.67	9.16	4.04
	Std Dev.	0.26	0.53	23.47	5.06	0.94
	Sample	2294	2294	2279	2279	2293
non-SA	Mean	0.48	2.30	41.00	8.89	4.00
	Std Dev.	0.26	0.54	24.31	4.80	0.93
	Sample	2125	2125	2112	2112	2124

Country	Statistics	Rural	Teacher absenteeism problem	School SES Average
BOTSWANA	Mean	0.49	1.69	0.10
	Std Dev.	0.50	0.60	0.62
	Sample	170	170	170
KENYA	Mean	0.67	1.57	-0.38
	Std Dev.	0.47	0.56	0.52
	Sample	185	184	185
LESOTHO	Mean	0.65	1.70	-0.31
	Std Dev.	0.48	0.60	0.40
	Sample	177	177	177
MALAWI	Mean	0.67	1.85	-0.47
	Std Dev.	0.47	0.58	0.51
	Sample	140	140	140
MAURITIUS	Mean	0.48	1.49	1.58
	Std Dev.	0.50	0.58	0.25
	Sample	153	153	153
MOZAMBIQUE	Mean	0.24	1.80	-0.31
	Std Dev.	0.43	0.51	0.56
	Sample	176	168	176
NAMIBIA	Mean	0.63	1.67	-0.13
	Std Dev.	0.48	0.62	0.82
	Sample	270	270	270
SEYCHELLES	Mean	0.16	1.62	1.40
	Std Dev.	0.38	0.50	0.18
	Sample	24	24	24
SWAZILAND	Mean	0.71	1.72	0.13
	Std Dev.	0.46	0.66	0.55
	Sample	168	168	168
TANZANIA	Mean	0.71	1.69	-0.52
(excl. Zanzibar)	Std Dev.	0.45	0.65	0.54
	Sample	181	181	181
UGANDA	Mean	0.80	2.05	-0.63
	Std Dev.	0.40	0.61	0.44
	Sample	163	163	163
ZAMBIA	Mean	0.47	1.67	-0.33
	Std Dev.	0.50	0.62	0.69
	Sample	173	169	173
ZANZIBAR	Mean	0.59	1.66	-0.23
	Std Dev.	0.49	0.61	0.63
	Sample	145	145	145
SOUTH AFRICA	Mean	0.43	1.72	0.58
	Std Dev.	0.50	0.60	0.80
	Sample	169	167	169
Total	Mean	0.56	1.71	-0.01
	Std Dev.	0.50	0.61	0.84
	Sample	2294	2279	2294
non-SA	Mean	0.57	1.71	-0.06
	Std Dev.	0.49	0.61	0.83
	Sample	2125	2112	2125

Table 2: Mean SES score by quintile and country

Quintile	1	2	3	4	5	Total
Botswana	-0.975	-0.487	-0.092	0.485	1.562	0.098
Kenya	-1.069	-0.791	-0.560	-0.223	0.735	-0.382
Lesotho	-1.034	-0.704	-0.431	-0.091	0.690	-0.315
Malawi	-1.143	-0.888	-0.650	-0.348	0.671	-0.473
Mauritius	0.835	1.453	1.701	1.886	2.083	1.583
Mozambique	-1.219	-0.891	-0.542	0.036	1.084	-0.307
Namibia	-1.186	-0.817	-0.439	0.294	1.489	-0.132
Seychelles	0.645	1.245	1.501	1.695	1.937	1.404
South Africa	-0.787	-0.128	0.596	1.287	1.948	0.582
Swaziland	-0.851	-0.410	-0.048	0.496	1.461	0.129
Tanzania	-1.200	-0.917	-0.690	-0.382	0.609	-0.516
Uganda	-1.279	-1.021	-0.801	-0.469	0.410	-0.633
Zambia	-1.241	-0.921	-0.585	0.037	1.051	-0.333
Zanzibar	-1.151	-0.833	-0.508	0.103	1.242	-0.231
Total	**-1.127**	**-0.750**	**-0.356**	**0.385**	**1.573**	**-0.055**
Non-SA	-1.115	-0.723	-0.304	0.487	1.621	-0.007

Table 3: Mean maths score by quintile and country

Quintile	1	2	3	4	5	Total
Botswana	490.57	499.01	510.21	508.05	556.62	490.57
Kenya	540.04	545.23	555.48	564.92	610.82	540.04
Lesotho	443.30	448.09	447.56	445.19	451.75	443.30
Malawi	422.44	426.81	435.32	433.23	446.84	422.44
Mauritius	519.33	564.17	587.07	619.93	639.55	519.33
Mozambique	526.43	524.93	530.59	530.15	538.18	526.43
Namibia	403.32	402.41	411.19	425.08	512.70	403.32
Seychelles	520.83	541.04	555.32	575.84	578.71	520.83
South Africa	441.84	444.74	454.18	491.47	596.97	441.84
Swaziland	505.56	511.29	510.92	513.43	541.48	505.56
Tanzania	484.29	511.44	528.54	527.62	560.14	484.29
Uganda	484.43	497.07	497.88	508.74	543.11	484.43
Zambia	414.20	425.74	435.61	434.33	466.10	414.20
Zanzibar	478.02	471.62	477.89	478.68	484.39	478.02
Total	**468.13**	**479.59**	**485.13**	**491.99**	**559.53**	**468.13**
Non-SA	469.09	480.13	488.01	494.35	557.25	469.09

Table 4: Mean pupil maths score by mother's education and country

Quintile	No schooling	Primary	Incom-plete Secondary	Matric	Post-School Training	Total
Botswana	489.6	507.6	524.9	517.2	537.9	510.9
Kenya	534.3	552.4	569.2	568.7	603.7	561.4
Lesotho	439.8	440.2	454.2	455.7	449.0	444.7
Malawi	430.1	425.9	435.2	449.6	453.6	430.6
Mauritius	504.3	551.2	605.5	617.9	659.2	582.1
Mozambique	529.9	531.1	527.7	522.8	546.2	530.5
Namibia	413.7	414.7	432.3	443.0	430.4	425.3
Seychelles	522.4	524.5	539.8	558.8	580.9	549.9
South Africa	436.5	452.2	476.3	506.4	513.6	475.4
Swaziland	505.0	514.0	512.7	521.3	524.0	514.5
Tanzania	486.4	523.6	508.3	546.4	573.4	521.2
Uganda	491.1	510.2	504.5	514.7	531.5	508.0
Zambia	418.6	425.2	429.5	451.7	459.0	433.6
Zanzibar	482.6	479.1	480.1	475.6	501.3	480.5
Total	**477.7**	**489.7**	**502.2**	**512.9**	**527.3**	**496.4**
Non-SA	480.3	491.3	504.5	513.6	528.6	497.9

Table 5: Mean pupil maths score by father's education and country

Quintile	No schooling	Primary	Incom-plete Secondary	Matric	Post-School Training	Total
Botswana	500.4	506.5	510.5	515.8	552.1	516.0
Kenya	532.4	545.9	549.6	566.7	597.9	563.6
Lesotho	436.2	441.4	446.2	454.7	466.7	445.2
Malawi	423.5	426.2	428.4	443.7	454.5	432.2
Mauritius	504.9	555.4	595.2	601.4	668.4	587.5
Mozambique	516.2	533.3	531.9	529.8	533.7	531.3
Namibia	411.8	414.9	430.2	434.7	467.0	433.3
Seychelles	564.6	532.8	559.8	546.6	605.8	563.9
South Africa	427.5	451.9	476.8	507.0	533.5	490.2
Swaziland	512.9	510.6	512.8	510.1	533.9	518.1
Tanzania	479.9	516.3	532.6	544.8	562.8	522.2
Uganda	478.9	502.0	509.4	510.9	532.8	508.3
Zambia	403.5	422.0	424.6	442.5	454.9	435.7
Zanzibar	466.1	480.2	483.1	480.7	489.6	478.8
Total	**469.4**	**490.2**	**500.7**	**503.9**	**532.0**	**500.2**
Non-SA	471.9	491.6	502.6	503.7	531.9	500.9

Table 6: Survey regressions

Dependent variable: Mathematics score				
	All	**SA only**	**Non-SA**	**All**
	(1)	**(2)**	**(3)**	**(4)**
Pupil level variables:				
SES	2.133	-15.742	2.06	-1.668
	(2.79)***	(4.45)***	(2.67)***	(2.08)**
SES-squared	10.869	17.738	10.913	11.756
	(16.94)***	(6.64)***	(16.99)***	(18.67)***
SA*SES	-19.307			
	(5.76)***			
SA*SES-squared	8.161			
	(3.04)***			
Parent education (index)	1.457	1.583	1.48	1.494
	(4.07)***	(1.12)	(4.01)***	(3.67)***
Mother has degree (1=yes)	6.566	6.485	6.176	8.11
	(2.65)***	(0.85)	(2.35)**	(2.99)***
Item index	4.045	4.496	3.957	5.39
	(11.37)***	(4.47)***	(10.40)***	(14.15)***
10+ books at home (1=yes)	13.612	17.372	13.237	19.207
	(11.90)***	(4.01)***	(11.17)***	(14.98)***
Age	-2.212	-1.668	-2.262	0.02
	(7.50)***	(1.17)	(7.52)***	(0.07)
Rural (1=yes)	-4.443	13.353	-4.939	-12.181
	(4.10)***	(2.46)**	(4.44)***	(12.26)***
Stays with parents (1=yes)	10.794	11.800	10.667	15.544
	(8.78)***	(2.33)**	(8.43)***	(11.16)***
Meal index	3.49	2.944	3.527	3.757
	(14.41)***	(3.17)***	(14.08)***	(14.14)***
Own textbook (1=yes)	4.008	13.113	3.634	2.104
	(3.85)***	(3.23)***	(3.37)***	(2.04)**
Male (1=yes)	9.994	0.261	10.638	11.253
	(10.43)***	(0.06)	(10.80)***	(10.44)***
Extra tuition (1=yes)	6.603	-11.847	7.632	23.764
	(6.31)***	(2.84)***	(7.05)***	(21.68)***
Days absent per month	-2.221	-0.974	-2.288	-1.963
	(10.85)***	(1.34)	(10.67)***	(9.03)***
Repeater	-19.12	-17.393	-19.111	-20.387
	(17.00)***	(3.85)***	(16.48)***	(15.85)***
Homework given once per week (1=yes)	5.849	-7.839	6.575	5.731
	(3.88)***	(1.15)	(4.25)***	(3.53)***
Homework given most days (1=yes)	5.726	-4.952	6.537	6.445
	(3.32)***	(0.70)	(3.69)***	(3.37)***
School/teacher level variables				
School facility index	1.898	5.512	1.89	1.589
	(13.20)***	(11.04)***	(13.12)***	(12.35)***
SA*School facility index	2.162			
	(5.98)***			
School homework index	7.938	44.272	7.671	18.336
	(7.23)***	(7.38)***	(6.91)***	(17.01)***
SA*School homework index	30.798			

	(6.77)***			
% repeaters in school	-18.56	-44.393	-18.566	-17.282
	(7.28)***	(3.53)***	(7.23)***	(8.01)***
SA*%repeaters in school	-48.28			
	(5.23)***			
School mean parent education index	4.93	7.774	4.648	1.828
	(7.43)***	(2.58)***	(6.79)***	(2.82)***
Pupil-teacher ratio	-0.1	1.285	-0.109	-0.144
	(5.28)***	(3.92)***	(5.75)***	(8.45)***
Teacher has degree (1=yes)	22.113	-2.173	22.271	3.53
	(9.31)***	(0.44)	(9.37)***	(1.72)*
SA*Teacher has degree	-18.581			
	(3.67)***			
Teacher has teaching training (1=yes)	9.722	-25.668	9.671	5.503
	(5.26)***	(1.78)*	(5.20)***	(2.91)***
Teacher experience (years)	0.166	-0.289	0.207	0.037
	(3.05)***	(0.90)	(3.72)***	(0.70)
Teacher absenteeism problem ((1=yes)	-5.935	-7.398	-6.093	-6.348
	(8.70)***	(2.26)**	(8.66)***	(9.30)***
Botswana	109.006		71.908	
	(8.78)***		(37.76)***	
Kenya	177.379		139.918	
	(14.12)***		(72.96)***	
Lesotho	74.273		37.145	
	(5.97)***		(21.59)***	
Malawi	73.448		36.806	
	(5.93)***		(17.06)***	
Mauritius	113.772		75.906	
	(8.85)***		(22.86)***	
Mozambique	165.747		128.451	
	(13.18)***		(66.95)***	
Namibia	36.961			
	(2.96)***			
Seychelles	94.484		57.509	
	(7.40)***		(16.87)***	
Swaziland	125.291		88.702	
	(10.07)***		(51.35)***	
Tanzania (excl. Zanzibar)	142.539		105.113	
	(11.48)***		(44.19)***	
Uganda	147.065		110.187	
	(11.74)***		(45.87)***	
Zambia	60.537		23.503	
	(4.88)***		(10.76)***	
Zanzibar	99.673		62.3	
	(8.07)***		(29.37)***	
Constant	302.659	241.489	341.897	361.228
	(22.58)***	(6.33)***	(48.10)***	(53.92)***
Observations	35 790	2 099	33 691	35 790
R-squared	0.38	0.52	0.37	0.22

t statistics in parentheses
* significant at 10%; ** significant at 5%; *** significant at 1%

Table 7: Mean actual and predicted Maths score by SES quintile, South Africa only

Quintile	Actual	Predicted	Actual minus predicted
1	441.84	472.42	-30.58
2	444.74	475.46	-30.72
3	454.18	492.55	-38.37
4	491.47	528.59	-37.12
5	596.97	591.02	5.95
Total	486.15	515.60	-29.45

Table 8: Hierarchical Linear Model for Mathematics Test Scores

		Coefficient	Standard error	t-value	Degrees of freedom	Significance
Model for intercept:						
Intercept	γ_{00}	420.752	12.817	32.83	153	0.000
Mean SES	γ_{01}	14.979	3.679	4.07	153	0.000
Model for SES slope:						
Intercept	γ_{70}	4.095	1.031	3.97	153	0.000
Mean SES	γ_{71}	2.380	0.715	3.33	153	0.001
Other fixed effects:						
Over12	β_1	-11.989	2.565	-4.67	2863	0.000
Male	β_2	1.916	2.571	0.75	2863	0.456
EnglishSometimes	β_3	12.316	3.793	3.25	2863	0.002
EnglishAlways	β_4	17.671	4.961	3.56	2863	0.001
Livedwithparents	β_5	10.644	3.162	3.37	2863	0.001
AbsentFeesUnpaid	β_6	-12.518	5.994	-2.09	2863	0.037
Boooks11plus	β_8	7.904	3.332	2.37	2863	0.018
Repeat Once	β_9	-11.279	3.025	-3.73	2863	0.000
Repeat Twice	β_{10}	-12.626	4.687	-2.69	2863	0.008
Repeat 3+ times	β_{11}	-20.574	4.862	-4.23	2863	0.000
Absentfromschool	β_{12}	-1.422	0.581	-2.45	2863	0.015
MotherMatric	β_{13}	6.252	3.266	1.91	2863	0.055
Free State	β_{14}	11.133	13.241	0.84	2863	0.401
Gauteng	β_{15}	69.453	17.362	4.00	2863	0.000
Kwazulu-Natal	β_{16}	67.251	16.423	4.10	2863	0.000
Limpopo	β_{17}	34.922	14.851	2.35	2863	0.019
Mpumalanga	β_{18}	21.661	13.772	1.57	2863	0.116
Northern Cape	β_{19}	38.639	13.887	2.78	2863	0.006
Eastern Cape	β_{20}	36.618	13.782	2.66	2863	0.008
Western Cape	β_{21}	90.146	19.868	4.54	2863	0.000
Random effects		**Standard deviation**	**Variance**	**Chi-square**	**Degrees of freedom**	**P-value**
Intercept	U_0	48.499	2352.126	828.542	153	0.0000
Mean-SES	U_7	6.765	45.769	208.530	153	0.0020
Level 1	R	62.257	3875.956			

Source: Van der Berg (2005)

Figure 1: Kernel density curves on maths score, South Africa vs. other countries

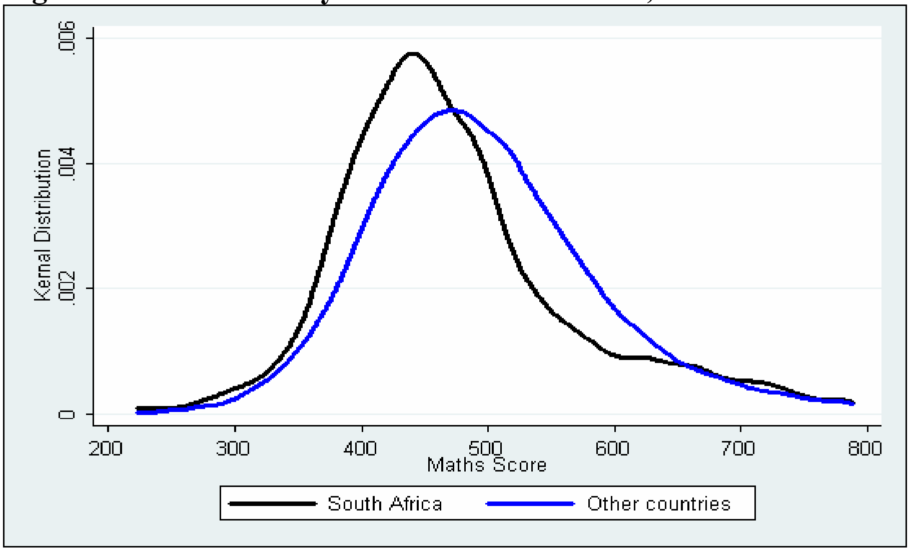

Figure 2: Lowess regression on pupil maths score, South Africa

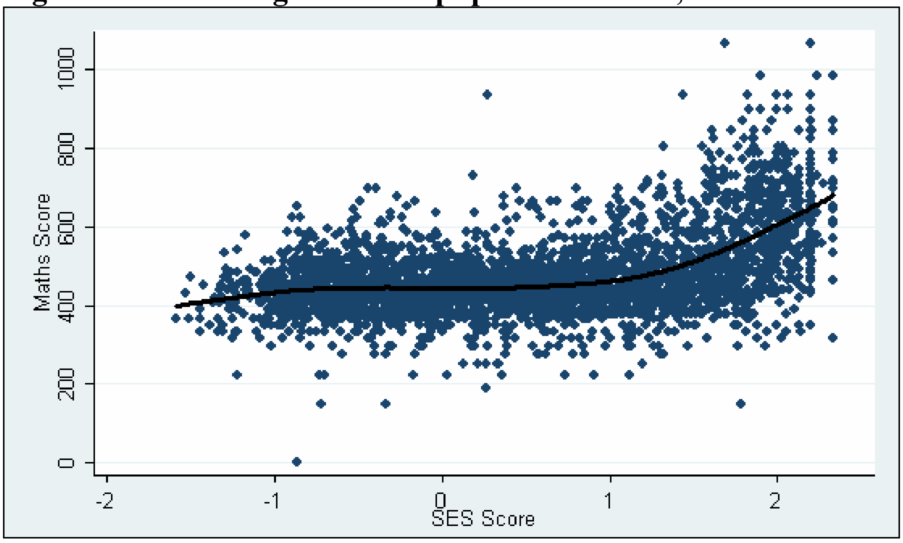

Figure 3: Lowess regression on pupil maths score, other countries

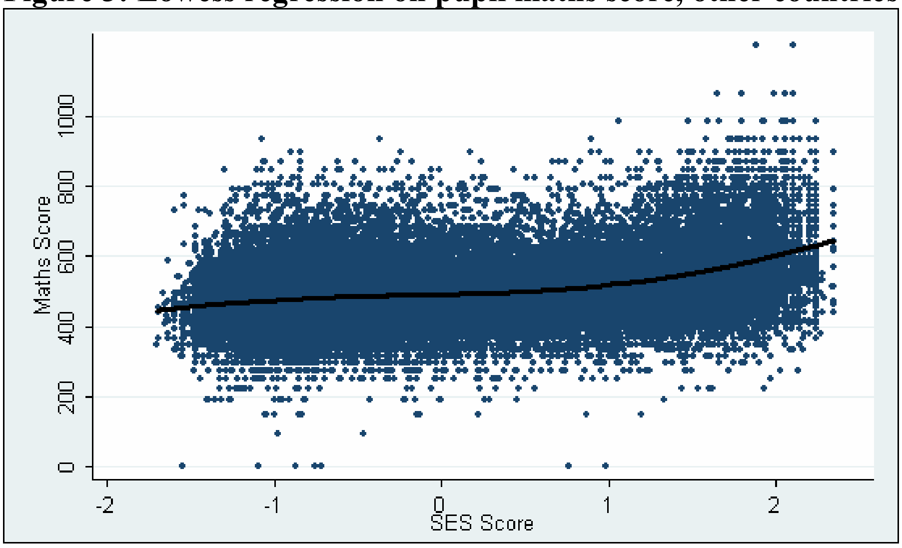

Figure 4: Lowess regression on schools' average maths score, South Africa

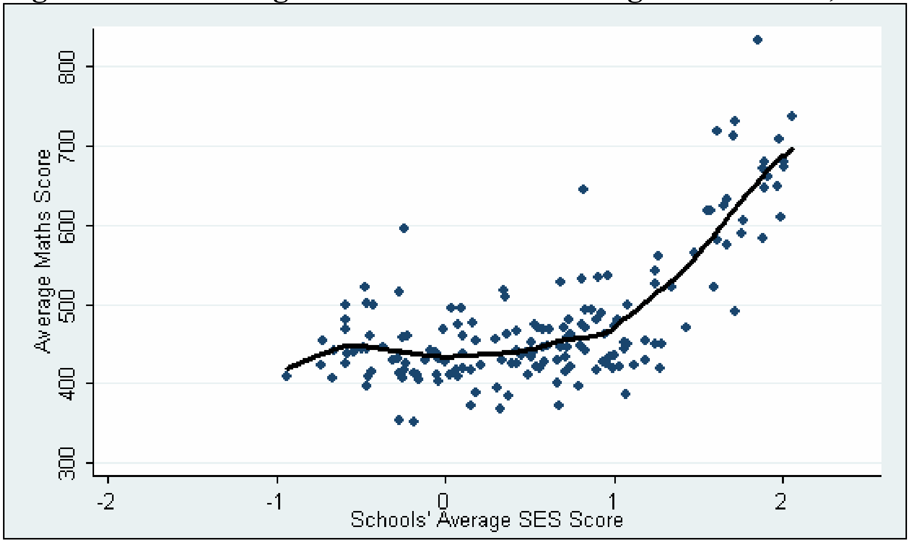

Figure 5: Lowess regression on schools' average maths score, other countries

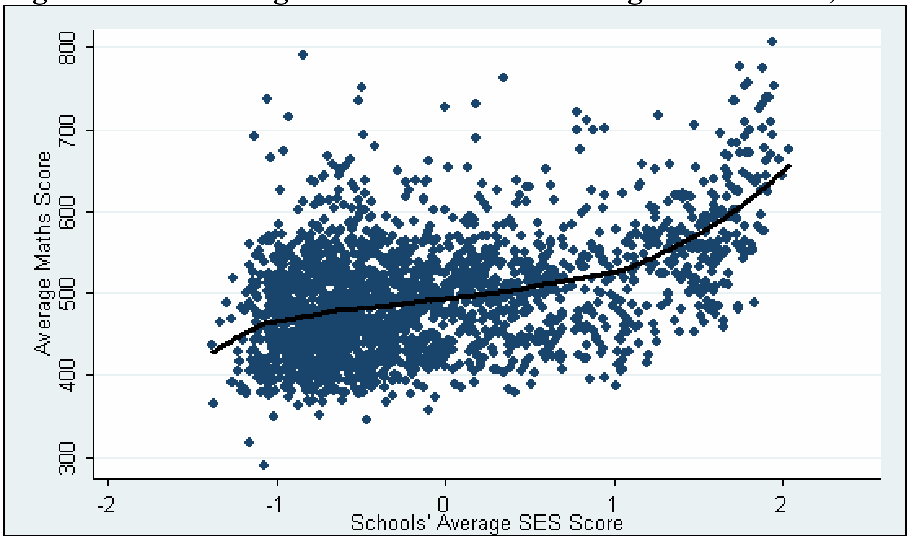

Figure 6: Lowess regression on pupil maths score, South Africa vs. other countries

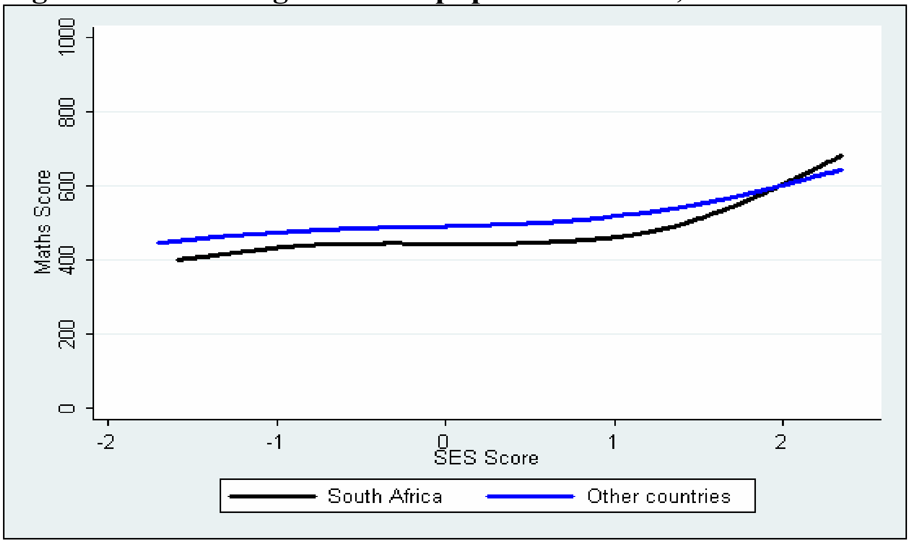

Figure 7: Lowess regression on schools' average maths score, South Africa vs. other countries

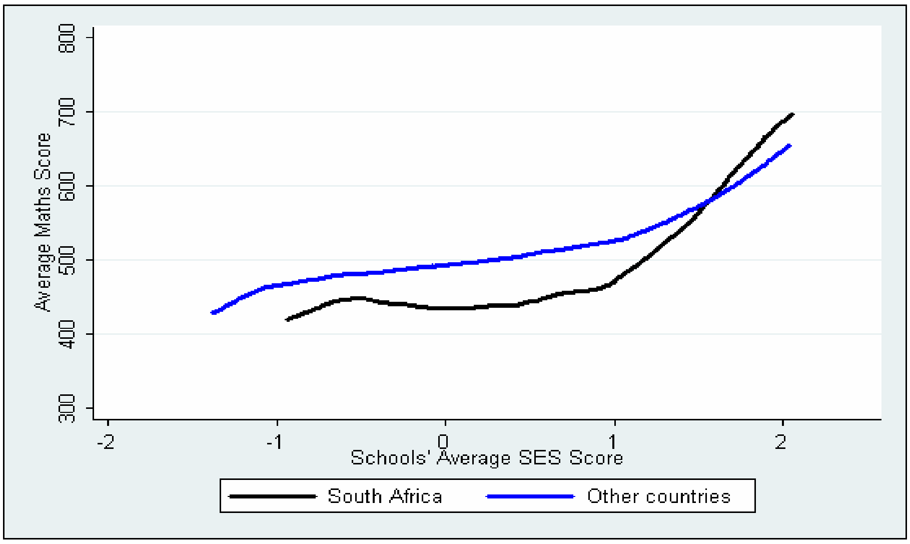

Figure 8: Mean SES and Maths score by country and quintile