Stephen Taylor

Department of Economics, University of Stellenbosch

Email: Stephen@sun.ac.za

Title:

Modelling indicators of effective school management in the South African education production function

Education production functions model cognitive skills as a function of personal characteristics that influence the learning efficiency of individuals and of various school-level inputs that influence learning. South African production function studies have typically faced two data limitations. Firstly, most datasets only capture student achievement at a single point in time, making it impossible to account for earlier influences on learning or innate ability. Secondly, most datasets are weak at measuring the quality of school management and teacher practices, which are theoretically expected to have important impacts on learning. Understanding what constitutes effective management is crucial in the light of an important finding in the literature: The impact of additional resources on student achievement is conditional upon how they are managed by schools.

The National School Effectiveness Study (NSES) provides an opportunity to go further in these two areas than was previously possible. This survey measured literacy and numeracy amongst a sample of South African learners in their third, fourth and fifth grade and collected information about management and teacher practice in astonishing detail by the standards of large sample surveys. Using education production function methods, this paper exploits these new data possibilities to identify specific indicators of effective school management and teacher practice that are associated with learning in South African schools. The paper also discusses how these findings might be translated into policies to improve school management.

JEL codes: I20,I21,I30,O15

1. Introduction

This paper represents an attempt to advance the South African education production function literature in two main ways, which flow from two distinct advantages of the National School Effectiveness Study (NSES) data.

The first contribution of this paper is facilitated by the rich collection of information about teaching and school management practices in South African primary schools provided by the NSES. Although aspects of school management are notoriously hard to measure, there are a number of reasons to expect this to be a key determinant of learning outcomes in South Africa. Quite apart from theoretical models of how management interacts with teaching and other contextual factors to influence learning, several empirical realities point to the likelihood that weak school management practices are driving learner underperformance.

Firstly, the finding that poor South African children are performing worse in reading and mathematics than equally poor children in the rest of Southern and East Africa despite generally better school resources (Van der Berg, 2007, Spaull, 2011) would suggest that it is not only poverty or inadequate school resources that underlie South Africa's weak performance, but also unobserved aspects of school functionality, management efficiency and teacher behaviour. This links to a second reason to suspect weak management practices: Learner performance has been largely unresponsive to considerable investments, especially within the historically disadvantaged part of the school system where education spending has been substantially increased since apartheid. It would appear that school resources do not *necessarily* make a difference but that the ability of schools to convert resources into outcomes is the crucial factor, and that this is where the policy attention is required (Van der Berg, 2008: 153).

The ability to convert resources into outcomes is one way to conceive of school efficiency. However, research based on nationally representative surveys has often been unable to illuminate the specific organisational features or teaching practices which promote greater school efficiency. Large-scale sample surveys of educational achievement are not always designed for a developing country context and therefore have typically not adequately captured the salient aspects of school management practice in South Africa. Moreover, when teachers and principals are asked in surveys about their behaviour and methods, they tend to give themselves a more favourable appraisal than reality would justify. Also, behaviour is likely to change upon observation. Consequently, when estimating statistical models, aspects of school practice such as time management may not appear to be significantly related to student achievement, even though such factors may indeed matter. School functionality or efficiency remains something of a "black box": resources flow into the box and differential outcomes emerge, yet little is known or can be proven about what occurs within the box to determine the outcomes.

Van der Berg and Burger (2002), in their analysis of achievement in the Western Cape province, found that approximately two-thirds of the variation in achievement between schools could be explained by socio-economic status, the racial composition of schools and a selection of teacher resource variables. They suggest that the efficiency of school management was probably an important omitted variable. Similarly, Crouch and Mabogoane (1998), combining the unexplained variation in their model with the effect of a dummy variable for historical education department (which they regard as capturing an efficiency dimension because socio-economic status was already controlled for in the model), estimated that approximately 50% of the variation in school performance was attributable to the unobserved feature of management efficiency. This large unexplained component that typically remains in South African production function studies suggest that unmeasured management practices may be a key driver of learner underperformance. As the next section describing the NSES data will argue, this collection of data contains more and better information on school management and teaching practices that has hitherto been collected in other large sample surveys of educational achievement such as TIMSS, PIRLS and SACMEQ.[1]

The second main contribution of this paper is that it models learning gains over time as opposed to learning achievement at a single point in time. This is due to the panel nature of the NSES data. Having gain scores offer several advantages over simple cross-sectional data. Prior influences such as innate ability, which cause omitted variable bias in cross section production functions, should now be controlled for in the baseline score. One can also examine how much learning has taken place during a given period. The use of gain scores, however, also introduces new complexities and challenges. The signal-to-noise ratio is diminished by the fact that there is now measurement error in both the pre-score and post-score. One can imagine that if the period between the tests is short enough the gain score will be likely to contain more noise than signal of actual learning. There is also a ceiling effect in so far as students with high pre-scores have less room to improve than those with low pre-scores. Finally, there is the statistical phenomenon of regression to the mean such that students with particularly low or high pre-scores will tend to score nearer to the sample mean in the post-test. These issues are discussed in greater detail in Section 3 and potential solutions are experimented with in the production function models.

2. Description of NSES data

Data for the National School Effectiveness Study (NSES) were collected between 2007 and 2009 on a nationally representative sample of schools in South Africa. The project was managed by JET Education Services and funded by the Royal Netherlands Embassy. Students in 266 schools

[1] TIMSS stands for Trends in International Maths and Science Study; PIRLS stands for Progress in International Reading Literacy Study; SACMEQ stands for Southern and East African Consortium for Monitoring Education Quality.

in eight of the nine provinces of South Africa were tested in literacy and numeracy in 2007 (grade 3), 2008 (grade 4) and 2009 (grade 5).[2]

The same individuals were tested in each year thus producing a panel dataset. The same tests were administered each year making the results comparable from one year to the next. This panel nature of the data is distinctly advantageous as it offers the potential to observe the amount of learning that occurs over time rather than a simple cross-sectional snapshot of achievement. In each year roughly 16 000 students were tested, although only 8383 students were matched across all three years. The reasons for this discrepancy include "dropping out" of the sample through grade repetition or moving away from a school, "dropping into" the survey (i.e. those who were not present in previous years joining the survey, for example, those repeating) and inefficiencies relating to the process of uniquely identifying and matching students across the waves of the survey.

Both the literacy and numeracy tests were administered in English to all students in all three years. The rationale behind this decision was to ensure comparability across the years, although it should be kept in mind that this places a handicap on non-English speakers, especially in grade 3 as policy stipulates that the switch from mother-tongue instruction to English should occur at the fourth grade. The literacy test consisted of 40 items and the numeracy test 53 items. Most items were short questions in which students either scored zero or one. However, several items in the literacy test required longer answers in which up to five marks were allocated. The scores used in this paper as outcome variables are percentage scores.

Table 1 reports several outcome measures of student achievement that can be calculated from the data. The mean percentage scores for literacy and numeracy in each wave of the survey are shown. The gain scores (in percentage points) are also shown from one year to the next and for the overall two-year gain. Note that the numbers in Table 1 are calculated using the sample of 8383 students that participated in all three waves of the survey.

Table 1: Mean weighted scores and gain scores in Literacy and Numeracy for all 3 waves

	Literacy	Numeracy
Grade 3 (2007)	20.15	29.38
Grade 4 (2008)	29.59	35.50
Grade 5 (2009)	37.73	47.04
Gain 2007 - 2008	9.43	6.12
Gain 2008 - 2009	8.14	11.54
2-year gain	17.57	17.66

[2] Unfortunately the project was blocked from surveying Gauteng due to other testing that was being administered in that province at the same time.

As Table 1 shows, the average achievement in literacy was lower than in numeracy, although both outcomes were at rather low levels given that the tests were comprised of items ranging from grade 1 to grade 4 levels. Nevertheless, there was clear improvement from one year to the next, with gains of about 17.5 percentage points for both literacy and numeracy over the two year period.

Figures 1 and 2 show the overall results by province for literacy and numeracy respectively. The lines show the percentage scores when calculated using only the sample of 8383 students that participated in all three waves (the panel sample). The columns show the percentage scores when calculated using the full cohort of students that participated in each year. The fact that the average scores were slightly higher for the panel sample than for the full cohorts indicates that those who remained in the sample throughout the period were a slightly stronger pool of students than those who for some reason participated in only one or two waves. This selection effect, however, does not appear strong enough to warrant suspicion that analysis based on only those captured in all 3 waves will be particularly biased.

Figure 1: Mean literacy scores in each year for the full cohorts and for the sample that was captured in all three years, by province

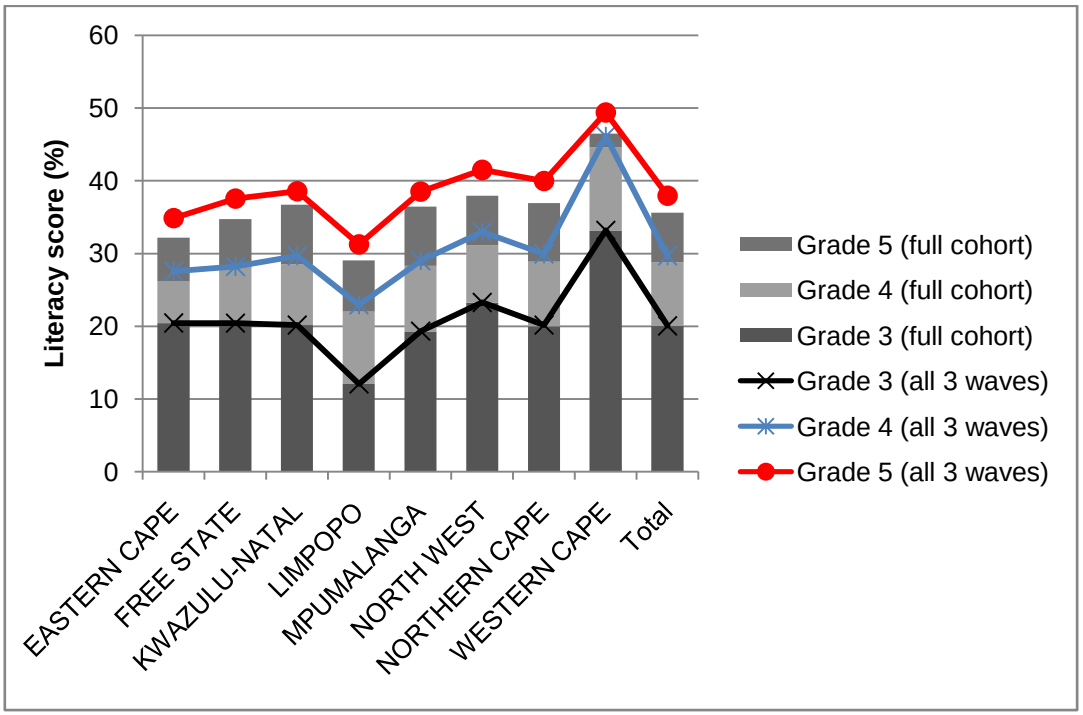
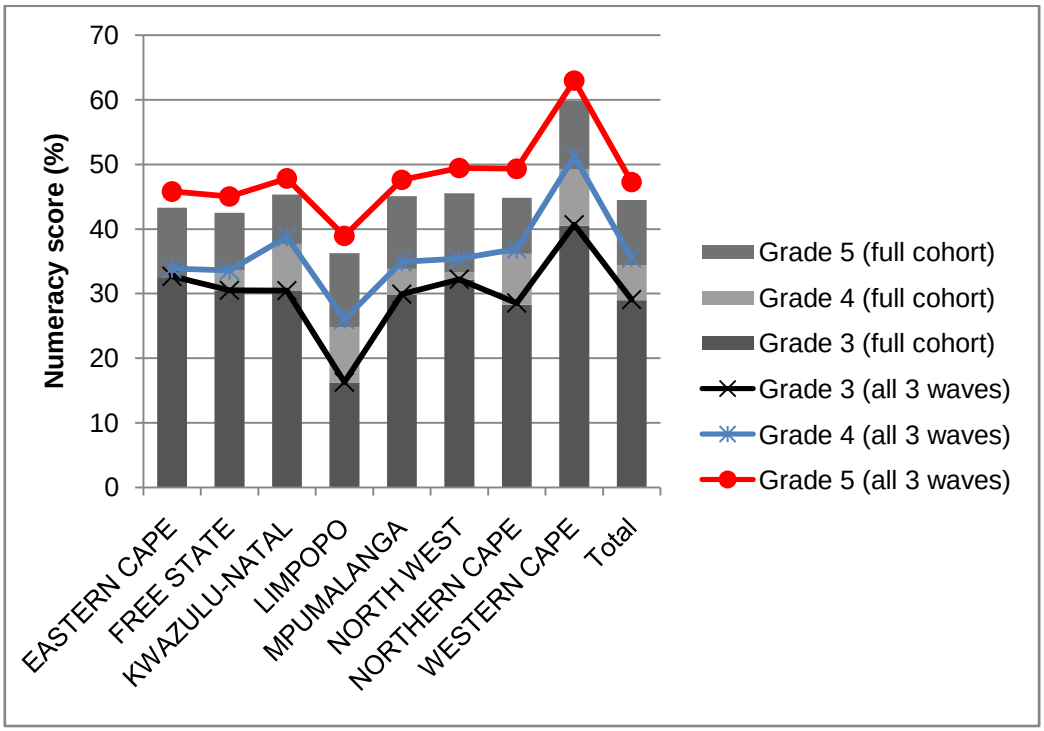


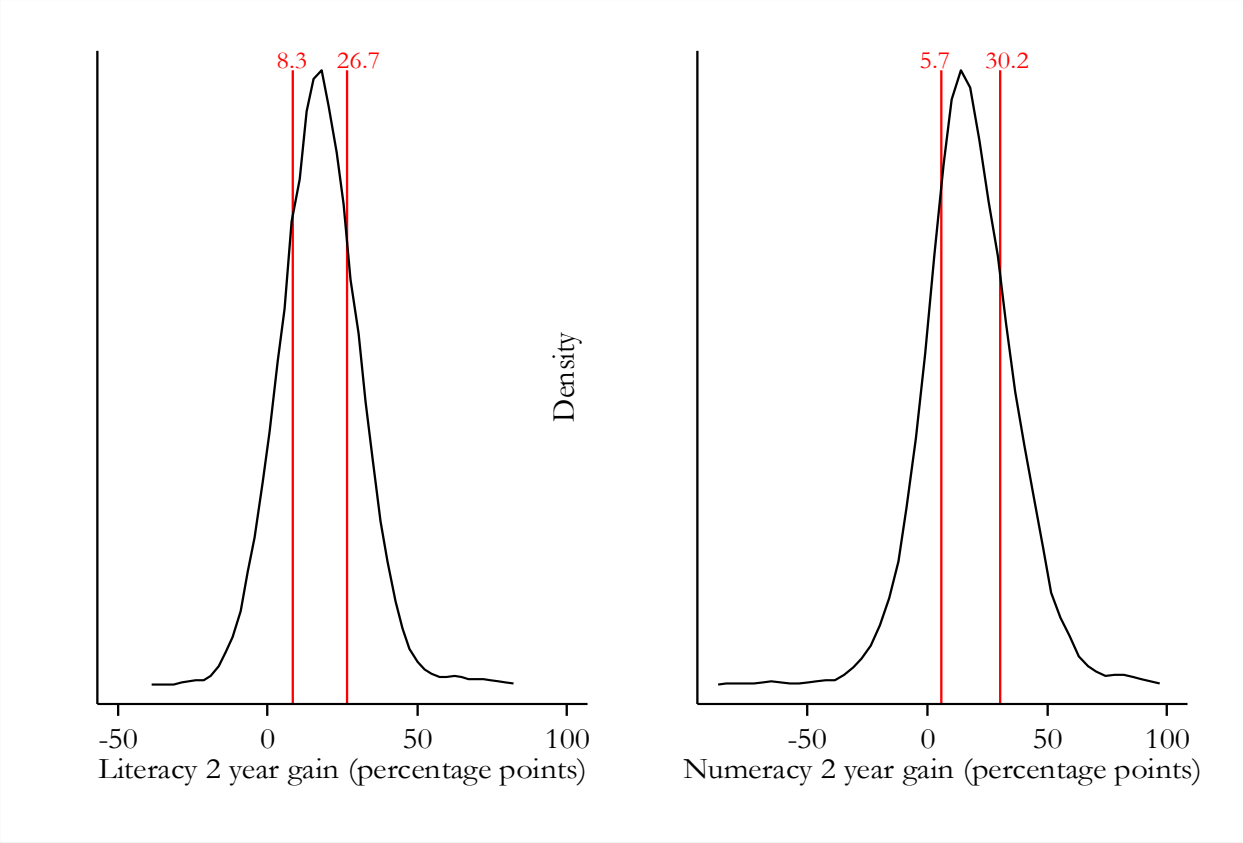
Figure 2: Mean numeracy scores in each year for the full cohorts and for the sample that was captured in all three years, by province



Figures 1 and 2 also show that Limpopo was the worst performing province in both literacy and numeracy, while the Western Cape was the best performing province. The substantial gain scores evident in the figures, even in low performing provinces, are an encouraging sign indicating that despite the low levels of achievement learning did occur over the two year period and that this was true for all provinces.

Figure 3 presents another way to view the gains, using kernel density curves. The distributions depict the proportion of students (along the vertical axis) achieving various gain scores in literacy and numeracy over the two years (along the horizontal axis). Additionally, vertical lines have been superimposed to indicate the position of the 25th and 75th percentiles within each distribution of gains. Several features of the graphs in Figure 3 are striking. The gain score at the 75th percentile was 26.7 percentage points in literacy and 30.2 percentage points in numeracy – fairly large gains indeed. The figure also draws attention to the large error component often present within gain scores due to measurement error at both the baseline test score and the post-test score. The very large positive gains and, especially, the large negative gains most likely represent measurement error rather than learning, or unlearning. Although measurement error may be a factor at the bottom end of the distribution of gains, it is still cause for concern that so many children exhibited very low or negative gains. 25% of the sample recorded gains of less than 8.3 percentage points in literacy and less than 5.7 percentage points in numeracy.

Figure 3: Kernel density curves of literacy and numeracy gains



Note: The red lines represent the 25th and 75th percentiles within each distribution of gain scores.

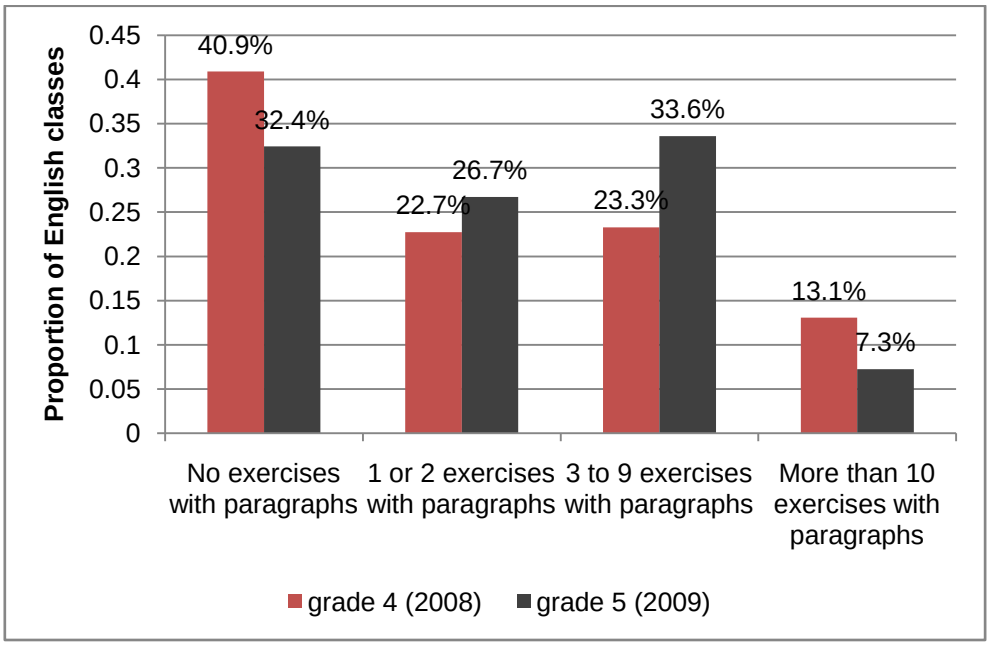
In addition to the testing, a wide variety of other information was collected through student questionnaires in 2007, 2008 and 2009, teacher questionnaires in 2008 and 2009 and school principal questionnaires in 2007, 2008 and 2009.[3] Consequently, the NSES boasts a rich collection of information regarding management and organisational practices within schools as well as teacher behaviour and practices. This derives from the sheer number of questions included in the principal and teacher instruments, the innovation of including short tests of content knowledge for teachers and an extensive review of student workbooks, which yielded several interesting indicators of curriculum coverage and the amount and type of work being done by children throughout the year. In order to avoid any bias caused by some teachers purposefully selecting the workbooks of more diligent students and other teachers selecting workbooks at random, teachers were asked to present the "best" student's workbook for inspection. Reviews of student workbooks were undertaken on this basis in 2008 and in 2009.

An example of the sort of information collected through the workbook review is the number of paragraph length writing exercises that were observed. Figure 4 shows, for grades 4 and 5, the proportions of classes in which various amounts of paragraph length writing was in evidence

[3] In addition, information on the ex-racial department of schools was imputed from the DoE's Master List of Schools.

according to learner workbooks. It is a matter for concern that in 2008 (grade 4) no evidence could be found of paragraph length writing in about 41% of classes, while in 2009 (grade 5) this figure was about 32%. In only a few classes was extended writing evidently a regular activity. It is therefore unsurprising that students answered longer items in the literacy test extremely poorly on average.

Figure 4: Frequency of paragraph length writing (grades 4 and 5) as a proportion of observations with valid responses[4]



3. Model assumptions

Consider that the observed literacy gain score (z_{it}) can be expressed as:

$$z_{it} = z_{it}^{*} + e_{it} \qquad (1)$$

[4] Note that there were 111 unspecified data entries for this variable in 2008 (grade 4). Therefore, this figure shows the percentage in each category out of the total number of valid data entries (i.e excluding unspecified cases from the denominator). The result is probably an overestimate of the amount of paragraph writing as one would expect that in most cases of unspecified data entries the learner workbooks were either incomplete or not shown to the fieldworker, which is already a sign of low teacher effectiveness. Therefore one should not conclude from the graph that there was a greater prevalence of classrooms in which more than 10 paragraph length exercises were conducted in grade 4 than in grade 5. Also, one can surmise that the proportion of classrooms in which no paragraph length writing was undertaken by learners was probably somewhat greater than 40.9% in grade 4.

where z_{it}^{*} is the true amount of learning that occurred between grade 3 and grade 5 and e_{it} is a mean-zero serially uncorrelated random error term.

Now, suppose that learning depends on prior knowledge and skills held at the start of the 2-year period of observation, collectively forming a student's learning proficiency (p_{it}^{*}), and all environmental, school and teacher influences during the period of observation (x_{it}^{*}). True learning (z_{it}^{*}) could therefore be expressed as follows:

$$z_{it}^{*} = x_{it}^{*} + p_{it}^{*} \qquad (2)$$

Note that student learning proficiency (p_{it}^{*}) would have developed as a function of innate ability together with the impacts of all environmental, school and teacher influences prior to the period of observation ($x_{i,t-1}^{*}$).

Now, the observed environmental, school and teacher influences during the period of observation (x_{it}) can be expressed as:

$$x_{it} = x_{it}^{*} + \varepsilon_{it} \qquad (3)$$

And therefore $x_{it}^{*} = x_{it} - \varepsilon_{it}$ (3.1)

Where ε_{it} is an error term capturing the effect of measurement error in the vector of observed environmental, school and teacher characteristics as well as the residual due to unobserved environmental influences.

The observed numeracy score in grade 3 (p_{it}) can be used as a proxy for the student's learning proficiency (p_{it}^{*}) and can similarly be expressed as follows:

$$p_{it} = p_{it}^{*} + u_{it} \qquad (4)$$

And therefore $p_{it}^{*} = p_{it} - u_{it}$ (4.1)

Where u_{it} is an error term capturing the extent to which numeracy scores in grade 3 deviate from true learning proficiency.

Now, substituting equations (3.1) and (4.1) into equation (2) gives the following:

$$z_{it}^{*} = (x_{it} - \varepsilon_{it}) + (p_{it} - u_{it}) \qquad (5)$$

And substituting equation (5) into equation (1) yields the estimable equation for observed literacy gain scores:

$$z_{it} = (x_{it} - \varepsilon_{it}) + (p_{it} - u_{it}) + e_{it} \qquad (6)$$

What can be done to deal with the three sources of error in the estimable equation? Error through unobserved environmental, school and teacher characteristics (ε_{it}) will always be present in education production function analysis due to the limitations of what can be observed with varying degrees of accuracy in surveys.

One potential solution to the problem of measurement error in grade 3 numeracy scores (p_{it}) is to use an instrumental variable (IV) approach. A valid instrument would need to be correlated with grade 3 numeracy scores (p_{it}) but not correlated with the measurement error (u_{it}). A suitable instrument for grade 3 numeracy scores is provided by grade 4 numeracy scores. There is of course measurement error in grade 4 scores also, but this error is uncorrelated with u_{it} so that it does not pose a problem.[5]

Observed literacy gains (z_{it}) deviate from true learning (z_{it}^{*}) through two processes both contributing to e_{it}. Firstly, the level of initial achievement determines the extent to which there is room for gains. Thus, it is not clear whether someone improving from a score of 10% to 20% learned more than someone improving from 80% to 88%. A second process is that of regression to the mean. Regression to the mean occurs because there is bound to be a degree of error whenever a student takes a test: Factors such as mood, health, concentration and so forth vary from one occasion to the next such that a student may have a "good" or a "bad" day. Because of this, low scoring students in a pre-test were more likely to have had a "bad" day than those who scored highly on the pre-test. As a result there will be a tendency for scores to regress somewhat to the mean score on the post-test. In order to account somewhat for these two processes one can include the literacy pre-score in the equation predicting the gain score.

When one accounts for both regression to the mean through the inclusion of the literacy pre-score (s_1) and for learning proficiency (p_{it}) one would expect the following:

[5] A similar strategy is followed by Ladd and Walsh (2002), although they estimate a 1-year gain score from grade 4 to grade 5 and use grade 3 scores to instrument for grade 4 scores.

$$E(z_{it} \mid p_{it}) = f(s_1) \text{ where } f'(s_1) < 0$$

That is to say that the expected gain score (z_{it}) for a student with a given learning proficiency (p_{it}) should be negatively related to the pre-score (s_1). I.e. if two students of identical ability were to score differently on the pre-test, one would expect a lower gain for the student who had initially scored higher. In an estimated regression one would therefore expect a negative coefficient on s_1 .

Similarly, one would expect that after accounting for regression to the mean, the gain score will be higher for students of greater proficiency:

$$E(z_{it} \mid s_1) = f(p_{it}) \text{ where } f'(p_{it}) > 0$$

In an estimated regression one would therefore expect a negative coefficient on p_{it} .

4. Results

The name "education production function" is appropriated from the industrial production function. This terminology has probably been responsible for some of the resistance to the education production function literature amongst some educationists, who correctly feel that educational outcomes cannot be generated through an optimal combination of inputs. A cautious interpretation of results is therefore required. For example, production functions do not prove causation and similarly, because school management and teaching quality are constructs comprising a wide spectrum of practices, the variables emerging as important may well be proxying for other characteristics. Nevertheless, education production functions hold potential for addressing important policy questions, especially in the case of South Africa. After accounting for the influence of SES, what school and teacher characteristics are associated with student achievement? Or slightly differently, what distinguishes better and worse-performing schools within poor communities?

Paul Glewwe (2002) warns that education production functions should be interpreted with caution due to numerous potential sources of bias. He therefore recommends an approach that sets out to gather as much evidence as possible and then make an overall judgement based on a number of estimations of production functions. This motivates the approach taken here in which several slightly different model specifications are estimated in order to address the various sources of bias identified in Section 3. Judgements about what school management and teaching practices are associated with learning will therefore take account of the sensitivity of results to model specification.

A description of all explanatory variables that were used in this analysis is presented in the Appendix. The summary statistics for these variables are also provided in the Appendix in Table 7. Numerous other variables were tested for inclusion in the models, but only those which were significantly related to gain scores in at least some model specifications were retained in the models presented here. Note that some variables appear related to achievement at one point in time but not to learning. Simple cross-section models using the NSES data are presented in Taylor (2011).

Table 2 shows the results from three survey regression models predicting the literacy gain from grade 3 to grade 5.[6] Model [A] predicts the gain score using a set of student, school and teacher characteristics, while in no way attempting to deal with any of the measurement error discussed in Section 3. Model [B] includes the literacy pre-score in the equation in order to deal with regression to the mean as well as grade 3 numeracy scores in order to account for learning proficiency, as described in Section 3. Model [C] is the same as Model [B] except that grade 4 numeracy scores are used to instrument for grade 3 numeracy scores.

The set of student characteristics that was associated with literacy gains was remarkably consistent across various model specifications. Student SES, although not significantly related to gains under the models, was included as this is a standard control in education production functions. Gender, age and exposure to English through speaking at home and through TV were consistently important student characteristics to include in the models.

Turning to school and teacher characteristics, it is interesting that the mean SES within each school (a measure that is known to be strongly correlated with educational achievement and school functionality in South Africa) was not significantly associated with gains in model [A] but after introducing the more nuanced model specifications in models [B] and [C] this did appear to be associated with gains.

Four variables that can be considered as indicators of management and teaching practice were consistently related to gains across all three models: teacher punctuality, whether the principal was absent on the day of the survey, paragraph writing and the number of literacy exercise observed in learner workbooks. Teacher punctuality and principal absenteeism reflect the quality of management and leadership present within a school, while ensuring that paragraph writing and other written exercises are undertaken is a reflection of teacher practice, but also of the instructional leadership within the school.

[6] Survey regression was used to account for complex sample design. The stratum variable was province (of which there were 8 due to the non-participation of Gauteng in the NSES), the Primary Sampling Unit (PSU) was the school (of which there were 266) and a person weight for each student, which differed only by province, was specified.

Table 2: Ways to model the literacy gain over two years

Explanatory variables	[A] no controls	[B] proficiency & rtm	[C] prof & rtm & IV[7]
Numeracy score (proficiency control)		0.18***	0.94***
Literacy score grade 3 (rtm control)		-0.69***	-1.59***
Student SES	0.14	0.28	0.49
Male	-1.79***	-2.74***	-3.39***
Young	-0.31	-1.26	2.45
Old	-3.79***	-3.88***	-4.48***
Speak English 1-3 times	1.53**	1.64***	1.27~
Speak English 4+	1.12	4.74***	4.37***
English on TV 4+	2.41***	3.96***	2.86***
Mean School SES	-0.38	1.93***	2.06*
Teacher punctuality good	1.83*	2.55***	3.14*
More than 2 English mark records	1.94	1.78*	0.06
Principal absent	-3.31**	-1.94*	-2.56~
Paragraph writing: none	-3.82***	-3.28***	-4.02**
Literacy exercises: more than 27	2.66**	3.22***	5.20***
Facilities index (2008)	0.04	0.24*	0.28
Monitoring through class visits	0.05	0.49	0.74
English teacher knowledge good	-0.26~	0.74	0.83
Constant	15.82***	15.25***	11.09***
R-squared	0.0883	0.3299	-
N	8383	8383	8383

~ p<0.10, * p<0.05, ** p<0.01, *** p<0.001

Similar indicators of teacher and management practice were related to gains in only some of the model specifications. These included a dummy for whether more than two English marks were recorded by teachers, an index summarising the presence and functionality of various school facilities, internal monitoring of curriculum coverage through class visits and the content knowledge of English teachers. It is interesting that these variables appeared more important when estimating gains on a restricted sample of students including only those attending historically black schools. The motivation for this restriction is discussed next.

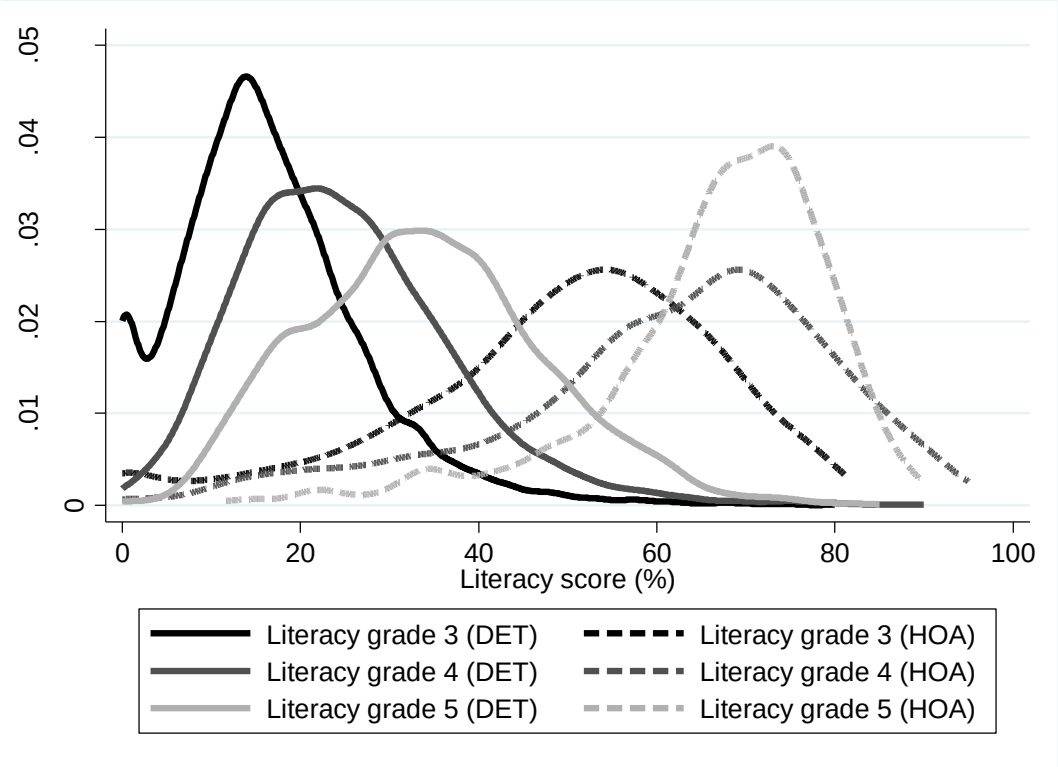
Many authors, social commentators and politicians have now used the idea of "two economies" or "two South Africa's" to describe the divided nature of various aspects of South African society. The education system in South Africa can similarly be characterized as consisting of two "sub-systems" that have very different historical backgrounds and continue to perform at different levels of effectiveness. The majority of South African children are located in the historically disadvantaged system, which continues to be disadvantaged through poverty and all the educationally detrimental factors that are associated with poverty. On average, children in

[7] A type of Hausman test was performed in order to test the validity of the instrument. This involved firstly regressing the grade 3 numeracy score on the grade 4 numeracy score (the instrumental variable) along with all the other regressors in the full model. The residuals from this regression were then included along with all the regressors in Model B in a second regression predicting the literacy gain score. The coefficient on the residual was statistically significant indicating that the instrument is adding information to the model.

these schools demonstrate low proficiency in reading, writing and numeracy. The second sub-system consists mainly of schools that historically served white children and produces educational achievement that is closer to the standards achieved in developed countries. This second system serves mainly white and Indian children, and increasingly serves black and coloured middle class children.

It is revealing to compare the distributions of achievement within each wave of the NSES for historically black schools with those for historically white schools. Figure 5 depicts these distributions for literacy. The three solid lines are for historically black schools (formerly administered by the Department of Education and Training – DET or by the various homelands) and the three broken lines for historically white schools (formerly administered by the House of Assemblies – HOA). For both groups of schools, the distribution of achievement improved with each year (shifting to the right). It is alarming, however, that the distribution for grade 5 students in historically black schools was still a considerably weaker distribution than that of grade 3 students in historically white schools. One can therefore conclude that by the fifth grade the educational backlog experienced in historically black schools is already equivalent to well over two years worth of learning.

Figure 5: Kernel Density curves of grades 3, 4 and 5 literacy by ex-department



This deficit despite more years of schooling may at least partly explain why the South African earnings function literature has found that white labour market participants enjoy higher returns to the same amount of education than black labour market participants (e.g. Burger and Jafta, 2006, Burger and Van der Berg, 2011). These studies suggest that the most probable

explanation for this result is that each additional year of education within the schools that black people typically attend does not lead to the same increase in productivity as that which is achieved during in each additional year within the schools typically attended by white people. This is indeed what is observed in Figure 5.

Not only is the distinction between the historically different parts of the school system useful for descriptive purposes, but there are important statistical and methodological reasons to analyse 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 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. In fact there is already a precedent for separate education production function analysis of the historically different systems in the literature (Gustafsson, 2007; Van der Berg, 2008).

Table 3 reports the results from the same three model specifications as in Table 2, but for the restricted sample of historically black schools only. The summary statistics for the explanatory variables for this sample is provided in Table 8 in the Appendix. The same set of student characteristics justified inclusion in these models. A key observation is that exposure to English through speaking at home and through TV remains important even when the model is restricted to historically black schools, suggesting that in the models for the full sample this result was not being driven by home language. Home exposure to English appears to be a valuable complement to having English as the language of instruction, which is supposed to be the case in grades 4 and 5.

As for the full sample, teacher punctuality, principal absenteeism, paragraph writing and the number of literacy exercises were consistently and strongly related to gain scores. This strengthens the case that these are reflecting important aspects of management and teacher practice that affect learning even after factors such as SES have been taken into account. It is interesting that the controls introduced in models [E] and [F] appear to bring out the importance of these school characteristics as well as several of the others which were ambiguously related to learning in model [D]. In particular, model [E] and [F] suggest that

English teacher content knowledge is significantly associated with learning gains. It is interesting that this did not emerge from the model for the full sample.

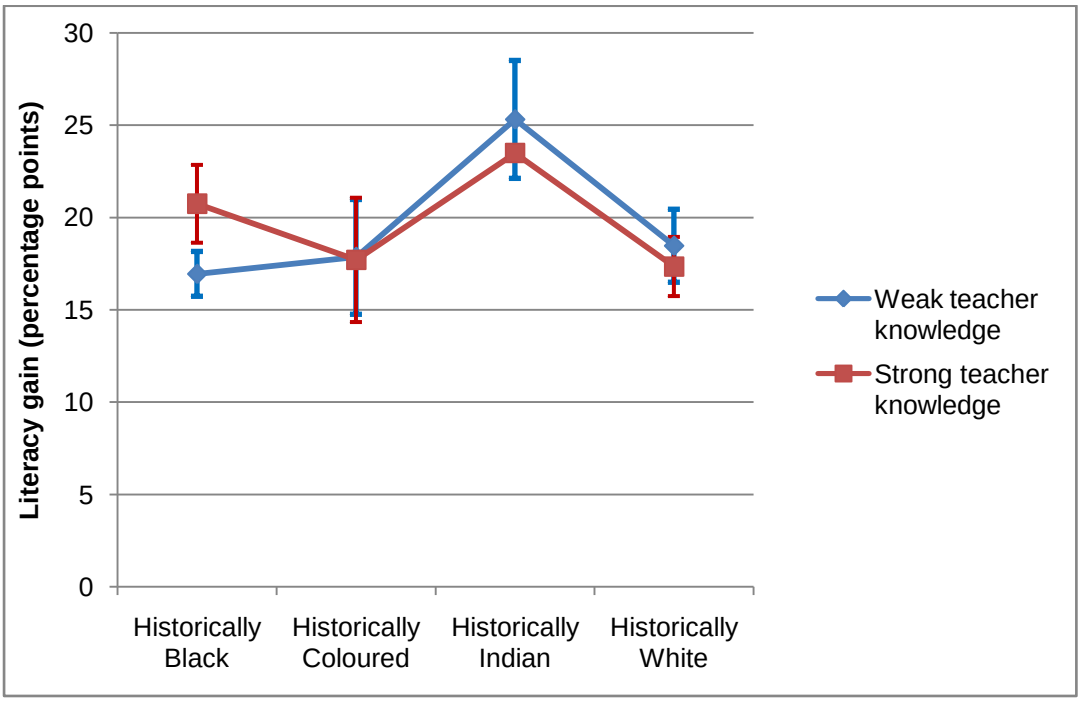
Table 3: Models restricted to historically black schools

Explanatory variables	[D] no controls	[E] proficiency & rtm	[F] prof & rtm & IV
Numeracy score (proficiency control)		0.17***	0.96***
Literacy score grade 3 (rtm control)		-0.70***	-1.57***
Student SES	0.24	0.32	0.66
Male	-2.17***	-2.79***	-3.23***
Young	-0.11	-0.61	3.02~
Old	-3.64***	-3.71***	-4.26***
Speak English 1-3 times	1.60***	1.75***	1.25**
Speak English 4+	1.53*	2.78***	3.07***
English on TV 4+	2.70***	3.87***	2.84***
Mean School SES	0.75*	1.44***	3.07***
Teacher punctuality good	2.58***	2.93***	3.65***
More than 2 English mark records	2.53***	2.44***	0.41
Principal absent	-3.63***	-2.53***	-1.64**
Paragraph writing: none	-3.24***	-3.34***	-3.75***
Literacy exercises: more than 27	2.93***	2.58***	5.30***
Facilities index (2008)	0.08~	0.27***	0.27***
Monitoring through class visits	0.50	-0.24	1.16**
English teacher knowledge good	0.41	1.02*	4.27***
Constant	12.24***	16.52***	7.24***
R-squared	0.120	0.319	-
N	6776	6776	6776

~ p<0.10, * p<0.05, ** p<0.01, *** p<0.001

The association of teacher content knowledge with learning is further considered at an unconditional level in Figure 6. In this case the measure of teacher content knowledge is calculated as a combination of the test performance of both maths and English teachers that taught each student in both 2008 and 2009. This measure therefore reflects the content knowledge of teachers in each school rather than that of an individual teacher. Such an aggregation is also motivated by the crude nature of the content knowledge tests. English teachers wrote a seven-item comprehension test and maths teachers wrote a five-item test. The maximum score on this derived measure is therefore 24 marks. The figure shows the mean literacy gain (as well as 95% confidence intervals) for students with teacher knowledge at least 18 out of 24 (good knowledge) versus those with weaker teachers. This is shown separately for each former education department. The figure confirms what the regression models suggested: teacher content knowledge appears to be related to student gains within the historically black schools but not within the rest of the system. This result requires further investigation but it may reflect that weak performance on the teacher test meant something different depending on which type of school the teacher was in, e.g. actual weak knowledge as opposed to low effort on the test due to other time pressures.

Figure 6: Mean literacy gain and 95% CI by teacher knowledge in both maths and language tests over both years (standard errors adjusted for clustering in schools & sampling weights)



5. Discussion

This paper has highlighted several indicators of effective school management and teacher practice that are associated with student achievement, even within the large historically disadvantaged and currently underperforming section of the school system. This constitutes an advance on earlier analyses which speculated about the importance of management efficiency but were limited in their ability to identify specific elements thereof.

Teacher knowledge was not consistently associated with achievement, although the evidence was stronger within the historically black sample of schools. Weak evidence was found that the quality of school facilities are associated with student achievement. As other studies have argued, more important than the mere presence of resources is how well they are managed. The results pertaining to variables that can be considered indicators of management effectiveness were clearer. Teacher punctuality, principal absenteeism, paragraph writing and the number of literacy exercises were consistently related to learning gains even after accounting for differences in previous student performance and SES.

The evidence suggests that effective schools offer thorough coverage of the curriculum. This should motivate policies aimed at teachers' professional development, firstly to ensure that they are technically able to teach all of the required elements, and secondly to provide them with the necessary time management skills so that they can deliver within the set academic timeframe. Policy should also ensure that support materials such as textbooks and workbooks are explicitly

designed to facilitate the extensive coverage of curriculum and exercises, making this easier for both teachers and students to implement.

The results point to the importance of instructional leadership in South African primary schools. The concept of "instructional leadership" is a rich education-specific framework for considering the role of principals as school managers. This concept emphasizes the role of principals as leader of curriculum coverage and teaching in the school. Hoadley, Christie and Ward (2009) find that the majority of principals do not regard the oversight of curriculum and teaching as their main task, but feel that the responsibility for this lies with subject heads. Perhaps as a consequence of this perception amongst principals, aspects of instructional leadership did not appear to take up the majority of their time but rather administrative duties and learner discipline (Hoadley *et al*, 2009: 381). Similarly, Reeves (2010: A-46) found that school management teams and heads of department in a Western Cape sample of mainly weakly-performing schools "do not appear to be playing a significant enough role in ensuring that grade-appropriate and sufficiently demanding reading, writing and numeracy activities and tasks are being done each day in class" in foundation phase classrooms. Towards the end of the second quarter of the school year, almost 60% of teachers had not been observed while teaching by a principal, member of the School Management Team or Head of Department (Reeves 2010: A-46). Applying the framework of instructional leadership to their empirical study of management in South African schools, Hoadley *et al* (2009) come to conclusions that align well to the evidence from the NSES discussed above: They find that successful instructional leadership occurs through ensuring curriculum coverage, good management of resources, facilitating parent support and structuring the school day effectively.

The indicators of good management identified in this research should, however, not be interpreted as more than exactly that: indicators that point to the characteristics typically exhibited by good managers, rather than levers to be manipulated by policy to achieve improved student outcomes. Command and control measures aimed at forcing teachers to follow best practices may well empty such practices of their value through introducing the perverse incentive to window-dress those practices at the expense of focusing on the central task of teaching. A better and indeed more ambitious route for policy would be to explore ways to attract, train and support better principals, and to replace those at the head of dysfunctional schools.

List of references

Burger, C., & Van der Berg, S. (2011). Modelling cognitive skills, ability and school quality to explain labour market earnings differentials. *Stellenbosch Economic Working Papers* (08/11).

Burger, R., & Jafta, R. (2006). Returns to race: Labour market discrimination in post-apartheid South Africa. *Stellenbosch Economic Working Papers* (04/06).

Crouch, L., & Mabogoane, T. (1998). When the residuals matter more than the coefficients: An educational perspective. *Studies in Economics and Econometrics , 22* (2), 1-14.

Glewwe, P. (2002). Schools and skills in developing countries. *Journal of Economic Literature , XL*, 436-482.

Gustafsson, M. (2007). Using the Hierarchical Linear Model to Understand School Production in South Africa. *South African Journal of Economics , 75* (1), 84-98.

Hoadley, U., Christie, P., & Ward, C. (2009). Managing to learn - Instructional leadership in South African secondary schools. *Research and Development Programme (TEP) Conference*, (pp. 138-151). Boksburg.

Reeves, C. (2010). Analysis of survey data. In S. Van der Berg, H. Meyer, C. Reeves, C. Van Wyk, U. Hoadley, M. Bot, et al., *Grade 3 Improvement Project: Main report and recommendations.* Report for the Western Cape Education Department and the Western Cape Treasury: University of Stellenbosch.

Spaull, N. (2011). A preliminary analysis of SACMEQ III South Africa. *Stellenbosch Economic Working Papers* (11/11).

Taylor, S. (2011). Uncovering indicators of effective school management in South Africa using the National School Effectiveness Study. *Stellenbosch Economic Working Papers* (10/11).

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

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

Van der Berg, S., & Burger, R. (2002). Education and socio-economic differentials: A study of school performance in the Western Cape. *South African Journal of Economics , 71* (3), 496-522.

Appendix: Description of variables used in multivariate analysis

Table 4: Student level variables (Student questionnaires of 2007, 2008 and 2009)

Variable name	Description
Student SES	Z-score index of socio-economic status: Min = 0, std dev = 1
Male	Dummy variable: gender is male; reference category is female
Young	Dummy variable: Younger than 10 years
Age 10	Dummy variable: Expected age at grade 4: 10 years
Old	Dummy variable: Older than 10 years
Speak English 1-3 times	Dummy variable: Student speaks English at home 1 to 3 times a week (reference category is never)
Speak English 4+	Dummy variable: Student speaks English at home more than 3 times a week (reference category is never)
English on TV 4+	Dummy variable: Student hears English on TV more than 3 times a week (reference category is 3 or fewer times)

Table 5: School level variables (Principal questionnaires of 2007, 2008 and 2009)

Variable name	Description
Mean School SES	Mean of student SES within each school: Min = 0, std dev = 1
Facilities index	Summative index capturing the presence and functionality of the following school facilities: running water, electricity, storerooms, toilets, administrative offices, box libraries and science kits.
Monitoring through class visits	Dummy variable: Internal curriculum monitoring takes place through class visits.
Principal absent	Dummy variable: The principal was absent in either 2008 or 2009 on the day of the survey.
Teacher punctuality good	Dummy variable: in all three years the principal maintained that teacher punctuality was not a serious problem in the school
More than 2 English mark records	Dummy variable: More than 2 mark records observed during the Principal instrument document review.

Table 6: Teacher level variables (Teacher questionnaire of 2008 and 2009)

Variable name	Description
Paragraph writing: none	Dummy variable: No evidence could be found in student workbooks of written exercises comprising paragraphs.
Literacy exercises: more than 27	Dummy variable: More than 27 exercises were counted in the “best” student’s English workbook.
English teacher knowledge good	Dummy variable: the sum of the teacher test scores of the student’s English teachers in grade 4 and grade 5 amounts to 13 or 14 out of a possible 14 marks.

Table 7: Summary statistics of variables used in multivariate analysis (full sample)

Variable	Mean	Standard deviation	Minimum	Maximum
Student SES	1.95	1.02	0	3.76
Male	0.47	0.50	0	1
Gender unspecified	0.02	0.13	0	1
Young	0.01	0.09	0	1
Old	0.12	0.32	0	1
Age unspecified	0.01	0.07	0	1
Speak English 1-3 times	0.38	0.49	0	1
Speak English 4+	0.12	0.33	0	1
Speak English unspecified	0.01	0.11	0	1
English on TV 4+	0.28	0.45	0	1
Mean School SES	2.13	0.96	0	3.98
Teacher punctuality good	0.48	0.50	0	1
Paragraph writing: none	0.29	0.46	0	1
Paragraph writing: unspecified	0.06	0.23	0	1
Literacy exercises: more than 27	0.59	0.49	0	1
Literacy exercises: unspecified	0.05	0.22	0	1
Monitoring through class visits	0.47	0.50	0	1
Monitoring unspecified	0.02	0.14	0	1
Facilities index	9.31	3.46	0	14
Principal absent	0.18	0.39	0	1
English teacher knowledge good	0.12	0.32	0	1
English teacher knowledge unspecified	0.30	0.46	0	1
More than 2 English mark records	0.79	0.41	0	1

Table 8: Summary statistics of variables used in multivariate analysis (historically black sample)

Variable	Mean	Standard deviation	Minimum	Maximum
Student SES	1.70	0.90	0	3.76
Male	0.46	0.50	0	1
Gender unspecified	0.02	0.14	0	1
Young	0.01	0.10	0	1
Old	0.14	0.35	0	1
Age unspecified	0.00	0.05	0	1
Speak English 1-3 times	0.38	0.49	0	1
Speak English 4+	0.07	0.26	0	1
Speak English unspecified	0.01	0.12	0	1
English on TV 4+	0.21	0.41	0	1
Mean School SES	1.85	0.79	0	3.74
Teacher punctuality good	0.46	0.50	0	1
Paragraph writing: none	0.34	0.47	0	1
Paragraph writing: unspecified	0.06	0.24	0	1
Literacy exercises: more than 27	0.53	0.50	0	1
Literacy exercises: unspecified	0.05	0.22	0	1
Monitoring through class visits	0.43	0.50	0	1
Monitoring unspecified	0.03	0.16	0	1
Facilities index	8.66	3.32	0	14
Principal absent	0.17	0.38	0	1
English teacher knowledge good	0.09	0.29	0	1
English teacher knowledge unspecified	0.30	0.46	0	1
More than 2 English mark records	0.77	0.42	0	1